HHS Public Access

Author manuscript

Clin Cancer Res. Author manuscript; available in PMC 2019 October 15.

Published in final edited form as:

Clin Cancer Res. 2018 October 15; 24(20): 5123–5132. doi:10.1158/1078-0432.CCR-18-0752.

Characterization of alternative splicing events in HPV negative head and neck squamous cell carcinoma identifies an oncogenic DOCK5 variant

Chao Liu[1,2], **Theresa Guo**[3], **Guorong Xu**[4], **Akihiro Sakai**[1], **Shuling Ren**[1,2], **Takahito Fukusumi**[1], **Mizuo Ando**[1], **Sayed Sadat**[1], **Yuki Saito**[1], **Zubair Khan**[3], **Kathleen M. Fisch**[4], and **Joseph Califano**[1,5]

[1]Moores Cancer Center, University of California San Diego, San Diego, CA, USA

[2]Department of Otolaryngology - Head and Neck Surgery, Xiangya Hospital, Central South University, Changsha, Hunan, China

[3]Department of Otolaryngology - Head and Neck Surgery, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA

[4]Center for Computational Biology and Bioinformatics, Department of Medicine, University of California San Diego, San Diego, CA, USA

[5]Division of Otolaryngology - Head and Neck Surgery, University of California San Diego, San Diego, CA, USA

Abstract

Purpose: Head and neck squamous cell carcinoma (HNSCC) is one of the most common cancers worldwide and alternative splicing is considered to play important roles in tumor progression. Our study is designed to identify alternative splicing events (ASEs) in HPV negative HNSCC.

Experimental Design: RNA sequencing data of 407 HPV negative HNSCC and 38 normal samples were obtained from The Cancer Genome Atlas (TCGA) and splice junctions were discovered using MapSplice. Outlier analysis was used to identify significant splicing junctions between HPV negative HNSCC and normal samples. To explore the functional role of the identified DOCK5 variant, we checked its expression with qRT-PCR in a separate primary tumor validation set and performed proliferation, migration and invasion assays.

Results: 580 significant splicing events were identified in HPV negative HNSCC and the most common type of splicing events was an alternative start site (33.3%). The prevalence of a given individual ASE among the tumor cohort ranged from 9.8% and 64.4%. Within the 407 HPV negative HNSCC samples in TCGA, the number of significant ASEs differentially expressed in each tumor ranged from 17 to 290. We identified a novel candidate oncogenic DOCK5 variant confirmed using qRT-PCR in a separate primary tumor validation set. Loss- and gain-of-function

Correspondence: Joseph Califano, MD, Division of Otolaryngology - Head and Neck Surgery, Department of Surgery, University of California San Diego, 3855 Health Science Drive, La Jolla, CA 92093, USA. Phone: 619-543-7895; jcalifano@ucsd.edu.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

This study has no potential conflict of interest that should be disclosed.

experiments indicated that DOCK5 variant promoted proliferation, migration and invasion of HPV negative HNSCC cells, and patients with higher expression of DOCK5 variant showed decreased overall survival.

Conclusion: Analysis of ASEs in HPV negative HNSCC identifies multiple alterations likely related to carcinogenesis, including an oncogenic DOCK5 variant.

Keywords

Head and neck squamous cell carcinoma; alternative splicing; HPV negative; DOCK5

INTRODUCTION

Head and neck cancer is the sixth most frequent malignant tumor worldwide and more than 90% of these cancers are head and neck squamous cell carcinomas (HNSCC). Despite great advances in surgery, radiotherapy, chemotherapy and even immunotherapy in the last decades, the 5-year survival rate for HNSCC is still relatively poor (1,2). Compared with HPV positive HNSCC, patients with HPV negative HNSCC have significantly worse prognosis (3,4). Therefore, there is an urgent need to advance our understanding of the underlying molecular mechanisms associated with HPV negative HNSCC carcinogenesis.

Alternative splicing events (ASEs) are a regulated process during gene expression that results in multiple mRNA and protein isoforms from a single gene. This process occurs in nearly all multi-exonic genes and increases the coding capacity of the human genome (5,6). Given that ASE plays an important role in the regulation of gene expression, aberrant splicing has thus been involved in a variety of human diseases including cancer (7). For instance, two splice variants of BCL2L1 have been described in cancer, BCL-X_L and BCL-X_s, which arise from an alternative 5' splicing site (8). The short isoform BCL-X_S has pro-apoptotic effects whereas the long isoform BCL-X_L is anti-apoptotic. In lymphoma and hepatocellular carcinoma, BCL-X_L isoform is predominant and protects cancer cells from p53-mediated apoptosis (9–11). With the development of next generation sequencing technologies and bioinformatics, more and more cancer-specific splicing patterns have been discovered. These splice variants could be used as hallmarks for cancer and potential therapeutic targets (12). In HNSCC, our previous study has identified several cancer related ASEs such as LAMA3 and DST variants using microarray analysis (13). And through RNA sequencing analysis, a novel functional splice variant of AKT3 was identified in HPV positive HNSCC, which could promote the proliferation of cancer cells (14). However, the pattern of ASEs in HPV negative HNSCC has yet to be elucidated.

In the past few years, genomic information related to various types of cancer has been annotated in databases such as The Cancer Genome Atlas (TCGA). These RNA sequencing data make it possible for systematic analysis of ASEs including novel splice variants in cancer. In our current study, we reanalyze RNA sequencing data from the TCGA HPV negative HNSCC cohort with MapSplice to detect splice variants and apply outlier analysis to identify tumor-specific ASEs. Compared with normal samples, we identified 580 significant alternative splicing alterations in HPV negative HNSCC, including a splice variant of DOCK5 with a unique end site. DOCK5 is a member of DOCK (dedicator of

cytokinesis) family, and members of this family act as guanine nucleotide exchange factors (GEFs) for small Rho family G proteins to regulate various physiologic process such as cell development, autoimmunity and bone homeostasis (15,16). However, the role of DOCK5 are not well understood yet. In this study, through loss- and gain-of-function experiments, the DOCK5 variant is confirmed to promote proliferation, migration and invasion in HPV negative HNSCC cells. These results provide valuable clues toward elucidating the function of DOCK5 variant on HPV negative HNSCC carcinogenesis.

MATERIALS AND METHODS

Patient samples

Primary HPV negative HNSCC tumor tissue samples (n=27) and normal mucosal samples from uvulopalatopharyngoplasty (UPPP) surgeries of non-cancer affected patients (n=17) were obtained from the cohort described previously (17). All of these tissue samples were collected from the Johns Hopkins Tissue Core under an Institutional Review Board approved protocol (#NA_00036235). Patient studies were conducted in accordance with the Declaration of Helsinki. Informed consents were obtained from all of the patients prior to participation in the study.

The Cancer Genome Atlas dataset

Raw RNA-seq data (fastq files) and clinical data of HPV negative HNSCC and normal samples were obtained from the TCGA Research Network (TCGA Provisional version updated in 2016, http://cancergenome.nih.gov/). These TCGA data included 407 HPV negative HNSCC and 38 normal tissues. RNA-Seq by Expectation Maximization (RSEM) normalized gene expression values for the same samples were downloaded from the Broad GDAC Firebrowse website (http://firebrowse.org/).

Identification of splice variants

The method used for identification of cancer specific splice variants was reported previously (14). Alignment of TCGA RNA sequencing data was conducted with MapSplice (18) to the GRCh37/hg19 genome assembly. Splice junction data from alignment was extracted for the following analysis. Expression values of junction were normalized as RPM (reads per million) and log transformed. The junctions were filtered if there was no difference in expression between any tumor and any normal samples, as well as if the junctions mapped to X, Y and MT chromosomes. The junctions were then mapped to known genes and exons based on GRCh37/hg19 genome assembly, and considered as putative splicing events if they were identified either as a skip (junction that skips a known exon), insertion (junction that starts or ends outside a known exon) or deletion (junction that starts or ends within a known exon). Expression values of these selected junctions were normalized by the RSEM values for the genes which were downloaded from TCGA. Outlier analysis was performed to identify the significant junctions between tumor and normal samples.

Integrative Genome Viewer validation

Putative significant junctions identified from outlier analysis were then visualized in Integrative Genome Viewer (IGV, Broad Institute) (19). BAM files of RNA sequencing data

were loaded into IGV, and reads coverage were visualized at start and end of each junction. Putative junctions were confirmed if the overall gene expression was observed in both normal and tumor tissue, and a unique novel splicing event was identifiable in tumor samples. Junctions were then categorized as either alternative start site, alternative end site, canonical skipping, insertion, deletion, intron retention or noncoding.

Cell culture and reagents

Human HPV negative HNSCC cell line BHY, UM-SCC17B and Detroit562 were obtained from the Gutkind Laboratory at the University of California San Diego, Moores Cancer Center and JHU011 cell line was obtained from Division of Head and Neck Cancer Research at the Johns Hopkins University. Cells were fingerprinted and confirmed using short tandem repeat analysis (20). BHY, UM-SCC17B and Detroit562 cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Sigma Aldrich, St. Louis, MO, USA) while JHU011 cells were cultured in RPMI-1640 medium (Sigma Aldrich), supplemented with 10% fetal bovine serum (FBS) and 1% Penicillin-Streptomycin. All cells were maintained at 37°C in an atmosphere of 5% CO_2 and 95% relative humidity.

siRNA transfection

BHY and JHU011 cell lines were transfected with siRNA reagents using Lipofectamine RNAiMAX (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instruction. Three unique custom designed siRNAs targeting the specific exon of DOCK5 variant were purchased from GE Dharmacon (Lafayette, CO, USA) using ON-TARGETplus (Supplemental Table S1). ON-TARGETplus SMART pool DOCK5 siRNA (L-018931–00-0005) was used for knockdown the overall DOCK5 gene expression, and a scrambled ON-TARGETplus Non-targeting pool siRNA (D-001810–10-20) was used as a negative control (NC), and the parental cancer cells without transfection were used as a blank control (Blank).

Stable transfection

The pLenti-C-mGFP-P2A-Puro empty vector was obtained from OriGene Technologies (Rockville, MD, USA), and gene of DOCK5 wild-type and variant were synthesized and cloned into the vector by GenScript, Inc. (Piscataway, NJ, USA). Lentiviral particles were prepared for empty vector, DOCK5 wild-type and DOCK5 variant using 293T cells as the packaging cells. UM-SCC17B and Detroit562 cells were infected with these lentiviruses and selected with 1 μg/ml Puromycin (InvivoGen, San Diego, CA).

Quantitative real-time PCR

To validate gene expression of DOCK5 wild-type (DOCK5 WT) and variant (DOCK5 Var), primers and probes sets were designed specifically to span the junction between the canonical two exons (DOCK5 WT) as well as the canonical exon and the unique tumor novel exon (DOCK5 Var) using PrimerQuest tools (Integrated DNA Technologies). The sequences of primer and probe were listed in Supplemental Table S1. Touchdown PCR was performed to identify appropriate length of PCR product for primers used. Quantitative RT-PCR (qRT-PCR) was used to determine the gene expression.

Total RNA was extracted from tissues and cells using RNeasy plus mini kit (Qiagen, Hilden, Germany), and reverse transcription was carried out with high-capacity cDNA reverse transcription kit (Thermo Fisher Scientific). *β-actin* was used as an internal control (Hs01060665_g1 TaqMan Gene Expression Assays, Thermo Fisher Scientific). PCR quantification was performed using the $2^{-\Delta\Delta CT}$ method.

Viability assay

3×10^3 cells per well were seeded into 96-well plates, and cell viabilities were measured at 24, 48, and 72 hours after transfection using Vita Blue Cell Viability Reagent (Bimake, Houston, TX, USA). After incubated for 1 hour at 37°C in the assay solution, fluorescence (Ex = 530–570 nm, Em = 590–620 nm) was measured using a microplate reader (BioTek, Winooski, VT, USA). All the experiments were repeated three or more times.

Colony formation assay

For colony formation assay, different groups of cells were seeded into 6-well plates (1×10^3 cells per well) and incubated for 2 weeks. Then the colonies were fixed with 4% paraformaldehyde and stained with crystal violet. Each experiment was repeated independently in triplicate.

Cell scratch assay

Cell migration ability was examined by cell scratch assay. Briefly, transfected cells were seeded on 6-well plates and incubated to almost full confluence. Scratching was performed with a 200 μl plastic pipette tip, and the cells were cultured in serum-free medium. The initial gap width (0 h) and the residual gap width at 24–96 hours after scratching were observed and photographed under the inverted microscope. The experiment was performed in triplicate.

Transwell migration and invasion assay

Transwell migration assay was conducted using 8 μm pore size Corning Transwell migration chambers and invasion assay was performed using Corning BioCoat Matrigel Invasion chambers (Corning Inc., Corning, NY, USA) according to the manufacturer's protocol. Briefly, cells (1×10^5 cells for migration; 2×10^5 cells for invasion) with serum-free medium were seeded to the upper chamber. After incubation for 48 h, non-migrating or invading cells on the surface of the upper chamber were removed with a cotton-tipped swab. The migrated or invaded cells on the lower side were fixed with paraformaldehyde, stained with 1% crystal violet, and then counted in five random fields under microscope.

Western blot analysis

Total cell proteins were extracted with RIPA lysis buffer, and the concentrations were measured using Protein Assay Kit (Bio-Rad, Hercules, CA, USA). Equal amounts of protein were separated on Mini-PROTEAN TGX gels (Bio-Rad) and transferred onto PVDF (polyvinylidene fluoride) membranes (Millipore, Billerica, MA, USA). After blocking with 5% BSA at room temperature for 30 min, the membranes were incubated with the relevant primary antibody at 4 °C overnight, followed by incubation with secondary antibodies for 1

h at room temperature. The primary antibody of p38, p-p38, Erk, p-Erk, MEK1/2 and p-MEK1/2 were all obtained from Cell Signaling Technology (1:1000, Danvers, MA, USA). Anti-GAPDH (1:10000, Cell Signaling Technology) was used as the loading control. Western blots were developed by ECL reagent (Pierce ECL Western Blot Substrate, Thermo Scientific).

Gene set analysis

Functional pathways associated with the oncogenic activities of DOCK5 variant were evaluated by gene set analysis. Hallmark gene sets were obtained from the Molecular Signatures Database (MSigDB, C2, Broad institute). Using the RNA sequencing data from TCGA HPV negative HNSCC samples, gene set enrichment analysis (GSEA) was performed (21) and data was compared between high DOCK5 variant expression and low DOCK5 variant expression samples. Benjamini–Hochberg correction was applied to P values from gene set analysis to correct for multiple comparisons and $P < 0.001$ was considered to be significant.

Statistical analysis

All assays were performed at least in triplicate and the results of the quantitative data represent mean ± standard deviation (SD) of three independent experiments. The statistical comparisons of two groups were determined with a two-sided unpaired Student's t test (for equal variances) or Mann–Whitney U test (for unequal variances) using SPSS software (version 23.0; SPSS Inc., Chicago, IL, USA). $P < 0.05$ was considered statistically significant.

RESULTS

Identification of significant ASEs in HPV negative HNSCC

For the TCGA RNA-seq data, a total of 4,660,670 putative junctions were identified by alignment with MapSplice. As shown in Fig. 1, after filtering junctions by multiple criteria, a total of 18,504 junctions were left for further analysis. Then these junctions were normalized by corresponding gene expression in each sample. Using an outlier analysis, 969 significant junctions were found differentially expressed in HPV negative HNSCC tissues compared with normal tissues. Each junction with statistically significance was then visualized in the IGV to validate the presence of a true ASE when comparing tumors and normal samples. Finally, 580 (59.9%) of significant junctions were confirmed after IGV visualization, involving 501 unique genes (Supplemental Table S2).

Characterization of significant ASEs

These 580 identified significant junctions were characterized as ASEs into the following categories: alternative start site (33.3%), alternative end site (20.2%), canonical variant or exon skipping (11.0%), insertion (17.1%), deletion (13.4%), intron retention (1.0%), and noncoding (4.0%) (Table 1). These ASEs were frequently expressed across multiple tumors, and the prevalence of a given individual ASE among the tumor cohort was between 9.8% and 64.4% of tumors harboring the same ASE (Supplemental Table S3). Within the 407 HPV negative HNSCC samples in TCGA, the number of ASEs identified in each tumor

ranged from 17 to 290 (median = 106, Supplemental Fig. S1). As shown in Supplemental Fig. S1, based on the number of ASEs in each tumor, the patients are divided into a high and low ASEs group. We found that there was no association between the number of ASEs identified within a tumor and the clinical parameters such as age ($P = 0.9213$), smoking status ($P = 0.7204$), alcohol consumption ($P = 0.0682$), clinical stage ($P = 0.1675$), pathologic stage ($P = 0.0728$) and lymph node metastasis ($P = 0.5397$). The patients with higher numbers of ASEs were more commonly occurred in male than female ($P = 0.0018$) (Supplemental Table S4). In addition, 10 of these identified significant junctions was also found in HPV positive HNSCC as described previously (14) (Supplemental Table S5), which indicated these 10 ASEs may play roles in both HPV negative and positive HNSCC.

DOCK5 variant is highly expressed in HPV negative HNSCC

Of these 580 identified significant junctions, junction chr8:25126401–25128306 within the gene DOCK5 was one of the most significant ASEs. Through IGV visualization, a novel exon with alternative end site was present in TCGA HPV negative HNSCC samples, while normal samples were harboring canonically splicing exons; and the amino acid length of DOCK5 variant was much shorter than DOCK5 WT (Fig. 2A). In TCGA data, most tumor samples have higher expression of DOCK5 variant compared with normal samples, and the total gene expression of DOCK5 was unchanged across tumor and normal samples (Fig. 2B). Compared with normal tissues, the expression of the DOCK5 variant was also up-regulated in tumors from diverse head and neck regions (Supplemental Fig. S2). To further confirm DOCK5 variant was specifically involved in HPV negative HNSCC samples, qRT-PCR was applied to compare the expression of DOCK5 variant in a validation set of 27 HPV negative HNSCC and 17 UPPP normal tissues. The results showed that DOCK5 variant was more highly expressed in tumor samples than UPPP normal tissues, while there were no significant changes in wide-type DOCK5 gene expression (Fig. 2C). Meanwhile, TCGA clinical data showed that higher expression of DOCK5 variant was associated with decreased overall survival in HPV negative HNSCC patients (Fig. 2D). To analyze the relationship between DOCK5 variant expression and gene mutation, we selected the significantly mutated genes identified with cBioPortal (MutSig q-value < 0.1) in the TCGA HNSCC data (22,23). As shown in Supplemental Fig. S3, among these 75 significantly mutated genes, the mutations or copy-number alterations of *TP53*, *NSD1* and *OR2M5* gene were more commonly occurred in patients with higher expression of DOCK5 variant ($P < 0.05$), and *DPPA2* gene mutation or copy-number alteration was more commonly occurred in patients with lower expression of DOCK5 variant ($P < 0.05$). We also found that there was no significant association between the expression of DOCK5 variant and the clinical parameters such as age, sex, alcohol consumption, clinical stage, pathologic stage and lymph node metastasis. But the patients with higher expression of DOCK5 variant were more commonly occurred in smokers than non-smokers ($P = 0.0358$) (Supplemental Table S6). These results indicate that DOCK5 variant overexpression may play a significant role in HPV negative HNSCC initiation and progression.

DOCK5 variant promotes proliferation of HPV negative HNSCC cells

To further clarify the role of DOCK5 variant in cancer, specific siRNAs and pooled siRNAs were designed to silence the expression of the tumor specific variant of DOCK5 with an

alternative end exon and overall DOCK5 gene, respectively. Meanwhile, DOCK5 variant or wild-type cDNA was cloned into an expression vector to generate stable cell lines with overexpression of DOCK5 variant or wild-type. Using sequence-specific primers and probes, expression of DOCK5 variant and wild-type was detected in a panel of 14 HPV negative HNSCC cell lines (Supplemental Fig. S4). From this data, BHY and JHU011 cell lines with the highest expression of DOCK5 variant were selected for loss-of-function assays, and UM-SCC17B and Detroit562 with the lowest expression of DOCK5 variant were selected for gain-of-function assays. qRT-PCR results demonstrated that 3 specific siRNAs targeting the DOCK5 variant successfully downregulated the expression of the DOCK5 variant, with only modest inhibition of the DOCK5 wild-type gene expression in both BHY and JHU011 cells (Fig. 3A). More specifically, in BHY cells, compared with the NC group, the expression of DOCK5 variant was decreased (92.7 ± 0.3) %, (96.3 ± 0.5) %, (95.3 ± 0.2) %, and the expression of DOCK5 wild-type was decreased only (27.9 ± 0.3) %, (36.1 ± 3.2) %, (25.3 ± 2.5) % by DOCK5 variant 1, 2 and 3, respectively (All $P < 0.01$). In JHU011 cells, the expression of DOCK5 variant was decreased (88.4 ± 1.2) %, (90.8 ± 3.1) %, (85.6 ± 0.3) %, and the expression of DOCK5 wild-type was decreased (44.6 ± 5.2) %, (45.3 ± 7.8) %, (32.2 ± 1.3) % by DOCK5 variant 1, 2 and 3, respectively (All $P < 0.01$). As expected, pooled siRNAs designed to target the whole DOCK5 gene decreased both the wild-type and variant expression of DOCK5, however, the ratio of DOCK5 variant expression to DOCK5 wild-type expression was significantly inhibited by DOCK5 variant siRNA, not by pooled siRNA (Fig. 3A).

Cell viability showed that over 3 days' transfection, significant growth inhibition was seen in both BHY and JHU011 cells after treatment with three independent ASE specific siRNAs (Fig. 3B). Notably, this growth inhibition was significant compared with both controls (Blank and NC group) as well as knockdown of overall DOCK5 gene expression (si Pool group), although cell growth was also inhibited in si Pool group compared with Blank and NC group. Meanwhile, a colony formation assay showed similar results, in which the colonies in DOCK5 variant knockdown group were much fewer and smaller than control groups in both cell lines (Fig. 3C). For the gain-of-function assay, in UM-SCC17B and Detroit562 cells, qRT-PCR results showed that the expression of DOCK5 variant or wild-type were successfully upregulated with transfection of lentivirus of DOCK5 variant or wild-type (Fig. 3D). In both UM-SCC17B and Detroit562 stably transfected cells, cell viability assay showed that overexpression of DOCK5 variant significantly increased cell growth but overexpression of DOCK5 wild-type showed no differences on cell growth (Fig. 3E). Colony formation assay revealed that the colonies in DOCK5 variant overexpression (Var OE) group were much more and bigger than empty vector (EV) and DOCK5 wild-type overexpression (WT OE) group (Fig. 3F). Taken together, these results demonstrate that the tumor specific DOCK5 variant can promote proliferation of HPV negative HNSCC cells.

DOCK5 variant promotes migration and invasion of HPV negative HNSCC cells

To study whether DOCK5 variant affects the metastatic ability of HPV negative HNSCC *in vitro*, cell scratch assay and Transwell migration and invasion assay were used to exam the changes of cell migration and invasion after transfection. Cell scratch assay revealed that although cell migration ability was inhibited by knockdown of overall DOCK5, it is much

more significantly in DOCK5 variant knockdown group (Fig. 4A), and overexpression of DOCK5 variant could increase cell migration ability (Fig. 4B). Transwell migration and invasion assay demonstrated that the number of migrating or invading cells was significantly reduced by knockdown of DOCK5 variant (Fig. 4C and Fig. 4D) and increased by overexpression of DOCK5 variant (Fig. 4E and Fig. 4F). These results reveal that DOCK5 variant enhances the migration and invasion ability of HPV negative HNSCC cells.

DOCK5 variant activates p38 and Erk MAPK pathway

Dysregulation within the MAPK pathway plays a critical role in HNSCC progression, so we investigated whether the DOCK5 variant was involved in MAPK pathway in HPV negative HNSCC. Western blot results showed that following knockdown the expression of DOCK5 variant in BHY and JHU011 cells, the expression levels of p-p38, p-Erk and p-MEK1/2 all decreased (Fig. 5A). And overexpression of DOCK5 variant in UM-SCC17B and Detroit562 cells increased the expression of p-p38, p-Erk and p-MEK1/2, while there were almost no changes in total p38, Erk and MEK1/2 (Fig. 5B). These results indicate that the DOCK5 variant can activate p38 and Erk MAPK pathways in HPV negative HNSCC.

Differentially expressed gene and pathway analysis

To better understand the mechanisms of DOCK5 variant in cancer, TCGA HPV negative HNSCC samples with the highest expression of DOCK5 variant (1/3 of total samples, n = 136) and lowest expression of DOCK5 variant (1/3 of total samples, n = 136) were selected to conduct a differentially expressed gene set and pathway analysis. Compared with the low expression of DOCK5 variant group, 1018 up-regulated genes and 1163 down-regulated genes were found in high expression of DOCK5 variant group, including the top up-regulated genes *DGUOK*, *RANBP1*, *SNRPG*, *TPRKB*, *HSPB11* and top down-regulated genes *DOPEY2*, *FYCO1*, *PBRM1*, *CCSER2*, *MAST3* (Supplemental Fig. S5A and Supplemental Table S7). Meanwhile, 301 up-regulated pathways and 682 down-regulated pathways were determined in high expression of DOCK5 variant group, including the top up-regulated pathways like “KEGG_PROTEIN_EXPORT”, “CHANG_CORE_SERUM_RESPONSE_UP” and “REACTOME_MRNA_DECAY_BY_3_TO_5_EXORIBONUCLEASE”, and top down-regulated pathways like “SHEDDEN_LUNG_CANCER_GOOD_SURVIVAL_A4”, “REACTOME_NOTCH_HLH_TRANSCRIPTION_PATHWAY” and “KEGG_GNRH_SIGNALING_PATHWAY” (Supplemental Fig. S5B and Supplemental Table S8).

DISCUSSION

Recently, with the rapid advancement of genome sequencing technologies and bioinformatics, there is an enormous resource of publicly available genome dataset of tumors of different subtypes. Among these databases, TCGA represents the most comprehensive integrative genomic analysis of cancer, which accelerates the thorough understanding of the molecular mechanisms of cancer (24,25). For HNSCC, TCGA has also identified various genes and pathways that are frequently mutated including *TP53*, *PIK3CA*, *NOTCH1*, *CDKN2A* and others, contributing to the development of new preventive

strategies, diagnostic methods and cancer therapies for HNSCC (22). Independent of genetic mutations, alternative splicing is another mechanism by which a single gene may generate multiple mRNAs and protein variants with different and even opposite functions (8,12). In our current study, utilizing the RNA sequencing data of TCGA, ASEs unique to HPV negative HNSCC were systematically characterized, which showed that alternative splicing could represent an important functional mechanism of carcinogenesis in HPV negative HNSCC.

While the landscape of alternative splicing variants in several types of tumor has been identified (26–30), it has not been previously reported in HPV negative HNSCC, which is proposed to be a distinct disease from HPV positive HNSCC with poorer prognosis. Meanwhile, most of currently existing methods to define differential ASE expression is based on comparing mean expression values between tumor and normal samples, which is insufficient for analysis of heterogeneous cancer sample populations (31–35). Outlier statistics was suggested to better capture significant events of heterogeneous tumors that may have similar mean values (36,37), and our previous studies also demonstrated that applying outlier analysis to microarray and RNA sequencing data was well suited for identifying ASEs in HNSCC (13,14,38). Therefore, in this study outlier analysis was applied to generate the profile of significant ASEs, and 580 ASEs were confirmed in HPV negative HNSCC. Although some of these identified ASEs may merely represent passenger alterations in RNA levels, many may be functionally active in HPV negative HNSCC, such as the DOCK5 variant described.

Alternative splicing is a ubiquitous regulatory mechanism that affects more than 95% of multi- exonic genes, and it has long been considered as an important mechanism for expansion of the eukaryotic proteome and play important roles in initiation and progression in both solid and liquid tumors (8,39,40). To date several types of alternative splicing have been described such as exon skipping, alternative start or end splice sites, intron retention and others (41). Kim et al pointed out that the distribution of the types of ASEs was different between cancerous and normal tissues, in which cancer cells showed less exon skipping, but more alternative start or end sites than normal cells (42). Our results also revealed that of these validated ASEs, the majority were alternative start sites (33.3%) and alternative end sites (20.2%). This is similar to HPV positive oropharyngeal cancer, in which alternative start site was also the most common type of ASE noted (14). One interesting finding was that patients with higher numbers of ASEs were more commonly occurred in men than women (Supplemental Table S4). Compared ASEs identified in HPV positive oropharyngeal cancer (14), we found 10 ASEs were included in our results (Supplemental Table S5), such as FBXO3, CYB561A3, GRHL3, NLRP1 variants and so on, which may play similar roles in both HPV negative and positive HNSCC. The small number of ASEs that overlap between HPV positive and HPV negative tumors further confirms that these two are distinct entities of head and neck cancer. In addition, it should be noted that some ASEs, like LAMA3 and DST variants that have been previously proved to be upregulated in HNSCC, were not identified in our results (13). These differences might be attributable to different number and clinical features of samples, including stringent outlier statistic, as well as differences between RNA sequencing-based discovery compared to array-based discovery used previously. To define these alternatives, when we relaxed the stringency of the outlier

approach by changing offsets in our outlier statistic approach, these variants were differentially present in tumor samples.

Utilizing described outlier statistics, a functionally relevant splice variant of DOCK5 was confirmed in HPV negative HNSCC. DOCK5 belonged to the DOCK family of GEFs, which consisted of 11 DOCK proteins in mammals (43). And DOCK1 was the founding member of the DOCK family, which was widely reported to be involved in cancer survival, migration and invasion (44–46). Though DOCK5 possessed the greatest similarity to DOCK1, it was one of the least studied members in the DOCK family, and DOCK5 was associated with mast cell degranulation (47), neutrophils and osteoclasts activity (48–51), obesity (52,53) and epithelial invasion (54,55). With regard to cancer, the role of DOCK5 has been poorly understood thus far. To our knowledge, this was the first study to reveal the function of DOCK5 and its variant on HNSCC progression. We found that DOCK5 variant was highly expressed in HPV negative HNSCC and patients with higher expression of DOCK5 variant showed decreased overall survival in the TCGA cohort. However, there was no significance between the expression of DOCK5 variant and disease free survival (Fig. 2D). Analysis of the relationship between DOCK5 variant and clinical parameters shows that DOCK5 expression is correlated with the smoking status, implying that smoking related comorbidity may be responsible for the worse overall survival in DOCK5 variant expressing tumors (Supplemental Table S6). The DOCK5 variant increased cell proliferation, migration and invasion in HPV negative HNSCC cells. Particularly, this effect was dependent on the expression ratio of DOCK5 variant compared with DOCK5 wild-type. Knockdown of the DOCK5 variant alone produced significant growth, migration, and invasion inhibition; however, the effect was weakened when knocking down the expression of whole DOCK5 gene including both variant and wild-type, and overexpression of DOCK5 wild-type almost had no effect on cell growth, migration and invasion. It should be mentioned that this study is limited by some factors. The DOCK5 variant and wild-type were not knocked down independently as the specific siRNAs targeting DOCK5 variant also had a little inhibition of the DOCK5 wild-type and pooled siRNA also inhibit the expression of DOCK5 variant. In the stably overexpressed cells, the expression of DOCK5 variant was much higher than the expression of the wild-type construct, which may partly explain that there were no obvious phenotype changes in DOCK5 wild-type overexpressed cells.

The DOCK family GEFs contained DHR-1 and DHR-2 domains, in which DHR-2 were GEF catalytic domain (56). Because of the alternative end exon, the amino acid sequence was much shorter and there was no canonical DHR-2 domain in DOCK5 variant, which indicated the potential function of DOCK5 variant may be independent of GEF domains. In this study, we found that DOCK5 variant could activate p38 and Erk MAPK pathway in HPV negative HNSCC. However, more functional studies are needed to define the potential mechanism of DOCK5 variant in regulating HPV negative HNSCC progression; the identified differentially expressed genes and pathways between higher and lower expression of DOCK5 variant patients in this study may provide some clues toward elucidating the mechanism.

In conclusion, we identified 580 tumor specific splice variant candidates within HPV negative HNSCC, including a splice variant of DOCK5, which can promote the

proliferation, migration and invasion of HPV negative HNSCC cells. These alternative splicing variants will provide a solid foundation for the future exploration of their potential roles in HPV negative HNSCC and the functionally relevant DOCK5 variant may be a potential therapeutic target for the treatment of HPV negative HNSCC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

GRANT SUPPORT

This work has been supported by National Institute of Dental and Craniofacial Research and National Institute of Health (R01 DE023347 to J. Califano) and the UC San Diego Clinical and Translational Research Institute Grant (UL1TR001442 to K. Fisch).

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66(1):7–30 doi 10.3322/caac.21332. [PubMed: 26742998]
2. Cohen EE, LaMonte SJ, Erb NL, Beckman KL, Sadeghi N, Hutcheson KA, et al. American Cancer Society Head and Neck Cancer Survivorship Care Guideline. CA Cancer J Clin 2016;66(3):203–39 doi 10.3322/caac.21343. [PubMed: 27002678]
3. Gillison ML, Chaturvedi AK, Anderson WF, Fakhry C. Epidemiology of Human Papillomavirus-Positive Head and Neck Squamous Cell Carcinoma. J Clin Oncol 2015;33(29):3235–42 doi 10.1200/JCO.2015.61.6995. [PubMed: 26351338]
4. Nevens D, Nuyts S. HPV-positive head and neck tumours, a distinct clinical entity. B-ENT 2015;11(2):81–7. [PubMed: 26563006]
5. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 2008;40(12):1413–5 doi 10.1038/ng.259. [PubMed: 18978789]
6. Modrek B, Lee C. A genomic view of alternative splicing. Nat Genet 2002;30(1):13–9 doi 10.1038/ng0102-13. [PubMed: 11753382]
7. Pajares MJ, Ezponda T, Catena R, Calvo A, Pio R, Montuenga LM. Alternative splicing: an emerging topic in molecular and clinical oncology. Lancet Oncol 2007;8(4):349–57 doi 10.1016/S1470-2045(07)70104-3. [PubMed: 17395108]
8. Chen J, Weiss WA. Alternative splicing in cancer: implications for biology and therapy. Oncogene 2015;34(1):1–14 doi 10.1038/onc.2013.570. [PubMed: 24441040]
9. Xerri L, Parc P, Brousset P, Schlaifer D, Hassoun J, Reed JC, et al. Predominant expression of the long isoform of Bcl-x (Bcl-xL) in human lymphomas. Br J Haematol 1996;92(4):900–6. [PubMed: 8616083]
10. Takehara T, Liu X, Fujimoto J, Friedman SL, Takahashi H. Expression and role of Bcl-xL in human hepatocellular carcinomas. Hepatology 2001;34(1):55–61 doi 10.1053/jhep.2001.25387. [PubMed: 11431734]
11. Schott AF, Apel IJ, Nunez G, Clarke MF. Bcl-XL protects cancer cells from p53-mediated apoptosis. Oncogene 1995;11(7):1389–94. [PubMed: 7478561]
12. Oltean S, Bates DO. Hallmarks of alternative splicing in cancer. Oncogene 2014;33(46):5311–8 doi 10.1038/onc.2013.533. [PubMed: 24336324]
13. Li R, Ochs MF, Ahn SM, Hennessey P, Tan M, Soudry E, et al. Expression microarray analysis reveals alternative splicing of LAMA3 and DST genes in head and neck squamous cell carcinoma. PLoS One 2014;9(3):e91263 doi 10.1371/journal.pone.0091263. [PubMed: 24675808]
14. Guo T, Sakai A, Afsari B, Considine M, Danilova L, Favorov AV, et al. A Novel Functional Splice Variant of AKT3 Defined by Analysis of Alternative Splice Expression in HPV-Positive

Oropharyngeal Cancers. Cancer Res 2017;77(19):5248–58 doi 10.1158/0008-5472.CAN-16-3106. [PubMed: 28733453]

15. Laurin M, Cote JF. Insights into the biological functions of Dock family guanine nucleotide exchange factors. Genes Dev 2014;28(6):533–47 doi 10.1101/gad.236349.113. [PubMed: 24637113]
16. Nishikimi A, Kukimoto-Niino M, Yokoyama S, Fukui Y. Immune regulatory functions of DOCK family proteins in health and disease. Exp Cell Res 2013;319(15):2343–9 doi 10.1016/j.yexcr.2013.07.024. [PubMed: 23911989]
17. Sun W, Gaykalova DA, Ochs MF, Mambo E, Arnaoutakis D, Liu Y, et al. Activation of the NOTCH pathway in head and neck cancer. Cancer Res 2014;74(4):1091–104 doi 10.1158/0008-5472.CAN-13-1259. [PubMed: 24351288]
18. Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res 2010;38(18):e178 doi 10.1093/nar/gkq622. [PubMed: 20802226]
19. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol 2011;29(1):24–6 doi 10.1038/nbt.1754. [PubMed: 21221095]
20. Martin D, Abba MC, Molinolo AA, Vitale-Cross L, Wang Z, Zaida M, et al. The head and neck cancer cell oncogenome: a platform for the development of precision molecular therapies. Oncotarget 2014;5(19):8906–23 doi 10.18632/oncotarget.2417. [PubMed: 25275298]
21. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102(43):15545–50 doi 10.1073/pnas.0506580102. [PubMed: 16199517]
22. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 2015;517(7536):576–82 doi 10.1038/nature14129. [PubMed: 25631445]
23. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013;6(269):pl110.1126/scisignal.2004088. [PubMed: 23550210]
24. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res 2016;44(8):e71 doi 10.1093/nar/gkv1507. [PubMed: 26704973]
25. Deng M, Bragelmann J, Schultze JL, Perner S. Web-TCGA: an online platform for integrated analysis of molecular cancer data sets. BMC Bioinformatics 2016;17:72 doi 10.1186/s12859-016-0917-9. [PubMed: 26852330]
26. Venables JP, Klinck R, Koh C, Gervais-Bird J, Bramard A, Inkel L, et al. Cancer-associated regulation of alternative splicing. Nat Struct Mol Biol 2009;16(6):670–6 doi 10.1038/nsmb.1608. [PubMed: 19448617]
27. Klinck R, Bramard A, Inkel L, Dufresne-Martin G, Gervais-Bird J, Madden R, et al. Multiple alternative splicing markers for ovarian cancer. Cancer Res 2008;68(3):657–63 doi 10.1158/0008-5472.CAN-07-2580. [PubMed: 18245464]
28. Venables JP, Klinck R, Bramard A, Inkel L, Dufresne-Martin G, Koh C, et al. Identification of alternative splicing markers for breast cancer. Cancer Res 2008;68(22):9525–31 doi 10.1158/0008-5472.CAN-08-1769. [PubMed: 19010929]
29. Shah TM, Patel AK, Bhatt VD, Tripathi AK, Shah S, Shankar V, et al. The landscape of alternative splicing in buccal mucosa squamous cell carcinoma. Oral Oncol 2013;49(6):604–10 doi 10.1016/j.oraloncology.2013.03.431. [PubMed: 23566772]
30. Danan-Gotthold M, Golan-Gerstl R, Eisenberg E, Meir K, Karni R, Levanon EY. Identification of recurrent regulated alternative splicing events across human solid tumors. Nucleic Acids Res 2015;43(10):5130–44 doi 10.1093/nar/gkv210. [PubMed: 25908786]
31. Hu Y, Huang Y, Du Y, Orellana CF, Singh D, Johnson AR, et al. DiffSplice: the genome-wide detection of differential splicing events with RNA-seq. Nucleic Acids Res 2013;41(2):e39 doi 10.1093/nar/gks1026. [PubMed: 23155066]

32. Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci U S A 2014;111(51):E5593–601 doi 10.1073/pnas.1419161111. [PubMed: 25480548]
33. Afsari B, Geman D, Fertig EJ. Learning dysregulated pathways in cancers from differential variability analysis. Cancer Inform 2014;13(Suppl 5):61–7 doi 10.4137/CIN.S14066. [PubMed: 25392694]
34. Hartley SW, Mullikin JC. Detection and visualization of differential splicing in RNA-Seq data with JunctionSeq. Nucleic Acids Res 2016;44(15):e127 doi 10.1093/nar/gkw501. [PubMed: 27257077]
35. Afsari B, Guo T, Considine M, Florea L, Kagohara LT, Stein-O'Brien GL, et al. Splice Expression Variation Analysis (SEVA) for Inter-tumor Heterogeneity of Gene Isoform Usage in Cancer. Bioinformatics 2018 doi 10.1093/bioinformatics/bty004.
36. Zhu Z, Ihle NT, Rejto PA, Zarrinkar PP. Outlier analysis of functional genomic profiles enriches for oncology targets and enables precision medicine. BMC Genomics 2016;17:455 doi 10.1186/s12864-016-2807-y. [PubMed: 27296290]
37. Mori K, Oura T, Noma H, Matsui S. Cancer outlier analysis based on mixture modeling of gene expression data. Comput Math Methods Med 2013;2013:693901 doi 10.1155/2013/693901. [PubMed: 23690879]
38. Gaykalova DA, Vatapalli R, Wei Y, Tsai HL, Wang H, Zhang C, et al. Outlier Analysis Defines Zinc Finger Gene Family DNA Methylation in Tumors and Saliva of Head and Neck Cancer Patients. PLoS One 2015;10(11):e0142148 doi 10.1371/journal.pone.0142148. [PubMed: 26544568]
39. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. Nature 2010;463(7280):457–63 doi 10.1038/nature08909. [PubMed: 20110989]
40. Singh B, Eyras E. The role of alternative splicing in cancer. Transcription 2017;8(2):91–8 doi 10.1080/21541264.2016.1268245. [PubMed: 28005460]
41. Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet 2010;11(5):345–55 doi 10.1038/nrg2776. [PubMed: 20376054]
42. Kim E, Goren A, Ast G. Insights into the connection between cancer and alternative splicing. Trends Genet 2008;24(1):7–10 doi 10.1016/j.tig.2007.10.001. [PubMed: 18054115]
43. Gadea G, Blangy A. Dock-family exchange factors in cell migration and disease. Eur J Cell Biol 2014;93(10–12):466–77 doi 10.1016/j.ejcb.2014.06.003. [PubMed: 25022758]
44. Tajiri H, Uruno T, Shirai T, Takaya D, Matsunaga S, Setoyama D, et al. Targeting Ras-Driven Cancer Cell Survival and Invasion through Selective Inhibition of DOCK1. Cell Rep 2017;19(5):969–80 doi 10.1016/j.celrep.2017.04.016. [PubMed: 28467910]
45. Laurin M, Huber J, Pelletier A, Houalla T, Park M, Fukui Y, et al. Rac-specific guanine nucleotide exchange factor DOCK1 is a critical regulator of HER2-mediated breast cancer metastasis. Proc Natl Acad Sci U S A 2013;110(18):7434–9 doi 10.1073/pnas.1213050110. [PubMed: 23592719]
46. Zhang B, Li H, Yin C, Sun X, Zheng S, Zhang C, et al. Dock1 promotes the mesenchymal transition of glioma and is modulated by MiR-31. Neuropathol Appl Neurobiol 2017;43(5):419–32 doi 10.1111/nan.12321. [PubMed: 26946516]
47. Ogawa K, Tanaka Y, Uruno T, Duan X, Harada Y, Sanematsu F, et al. DOCK5 functions as a key signaling adaptor that links FcepsilonRI signals to microtubule dynamics during mast cell degranulation. J Exp Med 2014;211(7):1407–19 doi 10.1084/jem.20131926. [PubMed: 24913231]
48. Watanabe M, Terasawa M, Miyano K, Yanagihara T, Uruno T, Sanematsu F, et al. DOCK2 and DOCK5 act additively in neutrophils to regulate chemotaxis, superoxide production, and extracellular trap formation. J Immunol 2014;193(11):5660–7 doi 10.4049/jimmunol.1400885. [PubMed: 25339677]
49. Touaitahuata H, Morel A, Urbach S, Mateos-Langerak J, de Rossi S, Blangy A. Tensin 3 is a new partner of Dock5 that controls osteoclast podosome organization and activity. J Cell Sci 2016;129(18):3449–61 doi 10.1242/jcs.184622. [PubMed: 27505886]
50. Vives V, Cres G, Richard C, Busson M, Ferrandez Y, Planson AG, et al. Pharmacological inhibition of Dock5 prevents osteolysis by affecting osteoclast podosome organization while preserving bone formation. Nat Commun 2015;6:6218 doi 10.1038/ncomms7218. [PubMed: 25645278]

51. Vives V, Laurin M, Cres G, Larrousse P, Morichaud Z, Noel D, et al. The Rac1 exchange factor Dock5 is essential for bone resorption by osteoclasts. J Bone Miner Res 2011;26(5):1099–110 doi 10.1002/jbmr.282. [PubMed: 21542010]

52. El-Sayed Moustafa JS, Eleftherohorinou H, de Smith AJ, Andersson-Assarsson JC, Alves AC, Hadjigeorgiou E, et al. Novel association approach for variable number tandem repeats (VNTRs) identifies DOCK5 as a susceptibility gene for severe obesity. Hum Mol Genet 2012;21(16):3727–38 doi 10.1093/hmg/dds187. [PubMed: 22595969]

53. Say YH. The association of insertions/deletions (INDELs) and variable number tandem repeats (VNTRs) with obesity and its related traits and complications. J Physiol Anthropol 2017;36(1):25 doi 10.1186/s40101-017-0142-x. [PubMed: 28615046]

54. Frank SR, Kollmann CP, van Lidth de Jeude JF, Thiagarajah JR, Engelholm LH, Frodin M, et al. The focal adhesion-associated proteins DOCK5 and GIT2 comprise a rheostat in control of epithelial invasion. Oncogene 2017;36(13):1816–28 doi 10.1038/onc.2016.345. [PubMed: 27669437]

55. Sanders MA, Ampasala D, Basson MD. DOCK5 and DOCK1 regulate Caco-2 intestinal epithelial cell spreading and migration on collagen IV. J Biol Chem 2009;284(1):27–35 doi 10.1074/jbc.M808010200. [PubMed: 19004829]

56. Namekata K, Kimura A, Kawamura K, Harada C, Harada T. Dock GEFs and their therapeutic potential: neuroprotection and axon regeneration. Prog Retin Eye Res 2014;43:1–16 doi 10.1016/j.preteyeres.2014.06.005. [PubMed: 25016980]

TRANSLATIONAL RELEVANCE

HPV negative HNSCC represents a distinct clinical entity from HPV positive HNSCC. Recently, alternative splicing has been shown to be closely related to tumor progression. However, the characterization of alternative splicing in HPV negative HNSCC is still largely undescribed. In this study, we identified 580 significant splicing events in HPV negative HNSCC and we identified a novel DOCK5 variant highly expressed in a separate primary tumor validation set using qRT-PCR. We demonstrated that the DOCK5 variant played an oncogenic role in HPV negative HNSCC, which could promote proliferation, migration and invasion in HPV negative HNSCC cell lines, and patients with higher expression of DOCK5 variant showed decreased overall survival. Identification of alternative splicing in HPV negative HNSCC allows for a promising avenue to identify novel therapeutic targets and expands the understanding of HNSCC carcinogenesis.

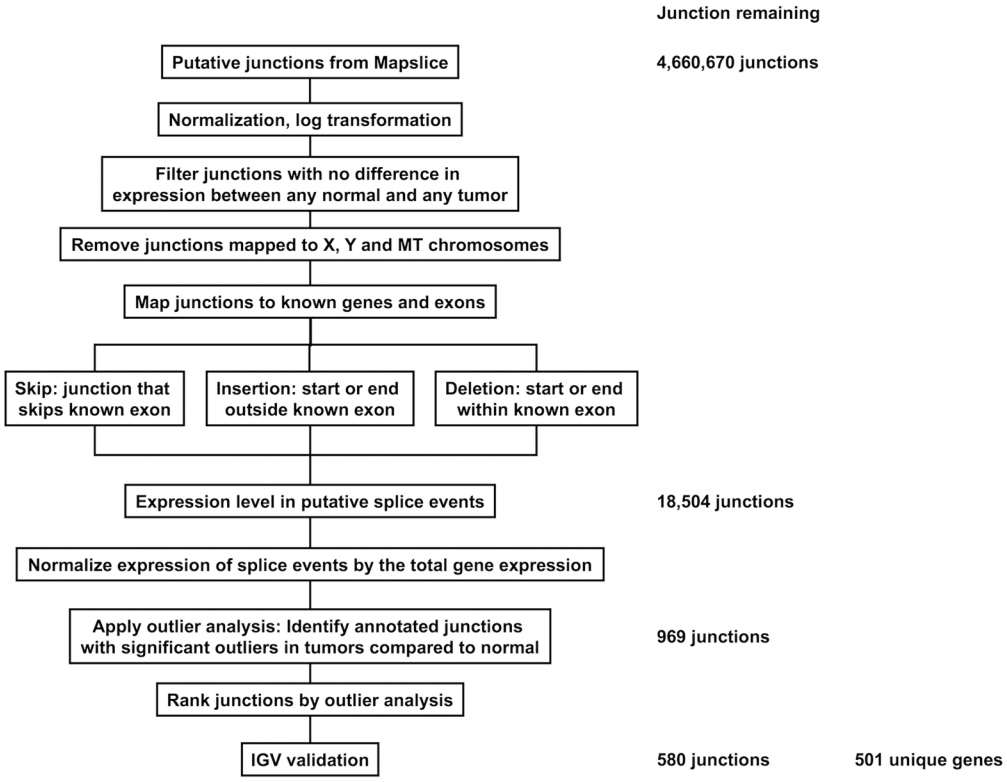

Figure 1. Pipeline for identification of significant ASEs through RNA-seq analysis.

Initially, from MapSplice output, 4,660,670 raw junctions, representing splice variant isoforms, were present within TCGA HPV negative HNSCC cohort. These junctions were normalized and then filtered to exclude junctions without variation or those on X, Y and mitochondrial (MT) chromosomes. Then potential splice variants were identified on the basis of possible alternative splicing patterns, and 18,504 potential candidates remained. Outlier statistics were applied to identify candidates with differential expression between tumors and normal samples, of these, 969 splice variants were identified. Finally, these junctions were validated by IGV visualization, and 580 junctions, which mapped to 501 unique genes were confirmed.

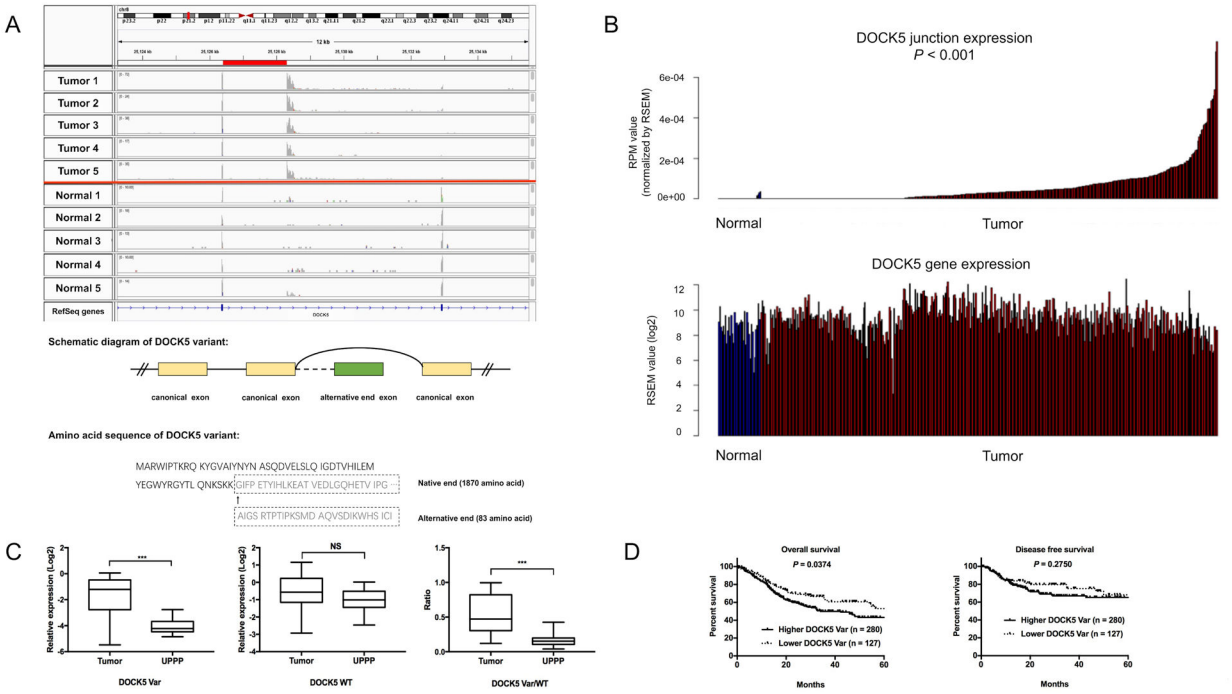

Figure 2. DOCK5 variant is highly expressed in HPV negative HNSCC

(A) IGV visualization of DOCK5 variant shows an alternative end site with expression unique to tumors (top), in which ten representative samples with reads coverage were shown; Schematic diagram of DOCK5 variant shows the alternative end site (middle); Amino acid sequence of DOCK5 variant shows the alternative end site, and the amino acid length of DOCK5 variant was much shorter than DOCK5 WT (bottom). (B) Unique DOCK5 junction expression and overall DOCK5 gene expression in TCGA HPV negative HNSCC data is shown. Expression of the DOCK5 splice junction is identified mostly in tumors. The corresponding overall DOCK5 gene expression (RSEM normalized) shows that these differences in DOCK5 splice variant expression are not fully explained by overall gene expression. Each vertical bar represents one sample sorted by expression level of DOCK5 variant (Normal samples: blue; Tumors: red) (C) Biologic validation of DOCK5 variant expression in primary tumor tissues. Quantitative PCR for identification of the DOCK5 variant shows significantly increased expression in tumors compared with normal UPPP tissues ($P < 0.001$), while wild-type DOCK5 gene expression is similar between tumors and normal tissues (NS), and the ratio of DOCK5 variant to wild-type is higher in in tumors than UPPP tissues ($P < 0.001$). (D) Log rank test shows that the patients with higher expression of DOCK5 variant have decreased overall survival ($P = 0.0374$), but no significant changes of disease free survival ($P = 0.2750$). The patients with no expression of DOCK5 variant (n = 127) are defined as lower expression of DOCK5 variant group. ***: $P < 0.001$, NS: not significant.

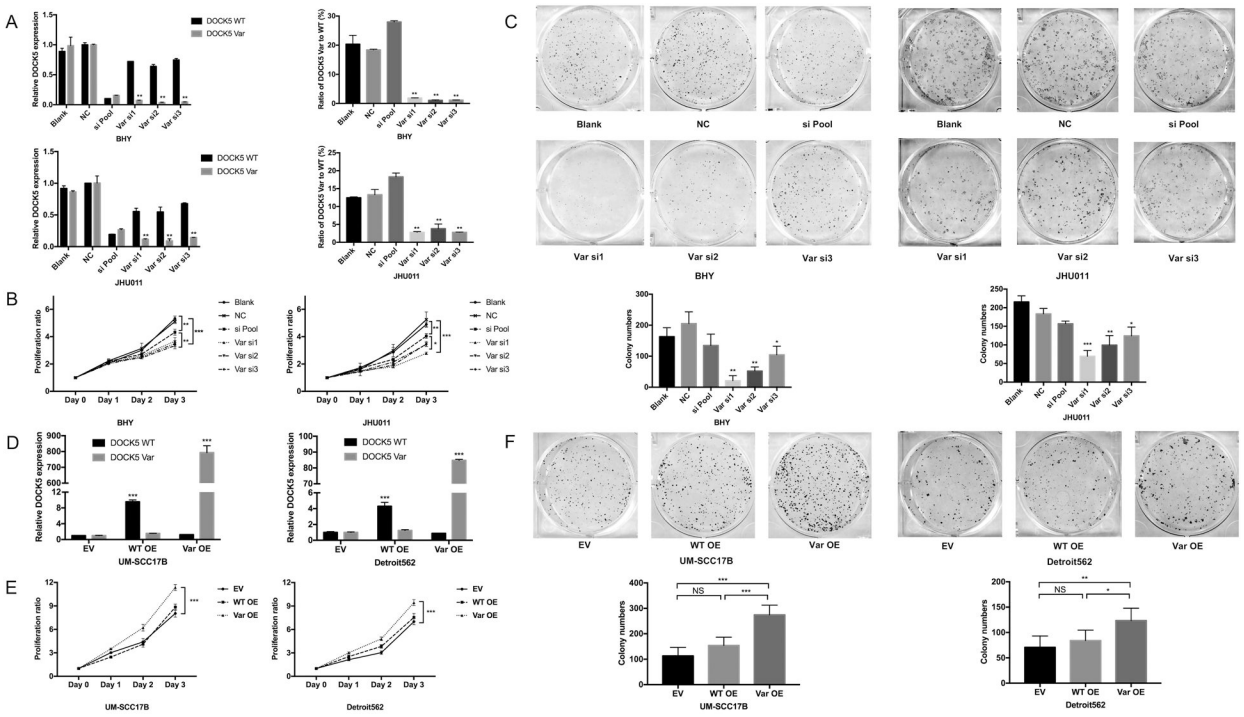

Figure 3. DOCK5 variant promotes proliferation of HPV negative HNSCC cells

(A) qRT-PCR results show the decreased expression of both DOCK5 wild-type (WT) and DOCK5 variant (Var) when treated with pooled siRNA for DOCK5. Three siRNAs specifically targeting DOCK5 variant show inhibition of the alternatively spliced DOCK5, with minimal inhibition of the wild-type gene. Specifically, the right panel shows the ratio of DOCK5 variant to wild-type expression, demonstrating that this ratio is significantly decreased by these three ASE siRNAs in both BHY and JHU011 cell lines. (B) Cell viability is measured after knockdown the expression of overall DOCK5 (si Pool) and DOCK5 variant (Var si1, si2, si3) in BHY and JHU011 cells. Proliferation ratio represented the cell numbers relative to day zero. Significant growth inhibition is seen with specific silencing of DOCK5 in both cell lines compared with silencing of the whole DOCK5 gene and control group (Blank, NC). (C) Colony formation assay shows the colonies in DOCK5 variant knockdown group are much fewer and smaller than control groups in BHY and JHU011 cells. (D) qRT-PCR results show that following transfection of lentivirus of DOCK5 variant or wild-type in UM-SCC17B and Detroit562 cells, the expression of DOCK5 variant or wild-type were successfully upregulated. (E) Cell viability assay shows significant growth increase by overexpression of DOCK5 variant, not DOCK5 wild-type. (F) Colony formation assay demonstrates that the colonies in DOCK5 variant overexpression group are more and bigger than empty vector groups and DOCK5 WT overexpression group in UM-SCC17B and Detroit562 cells. ***: $P < 0.001$, **: $P < 0.01$, *: $P < 0.05$, NS: not significant.

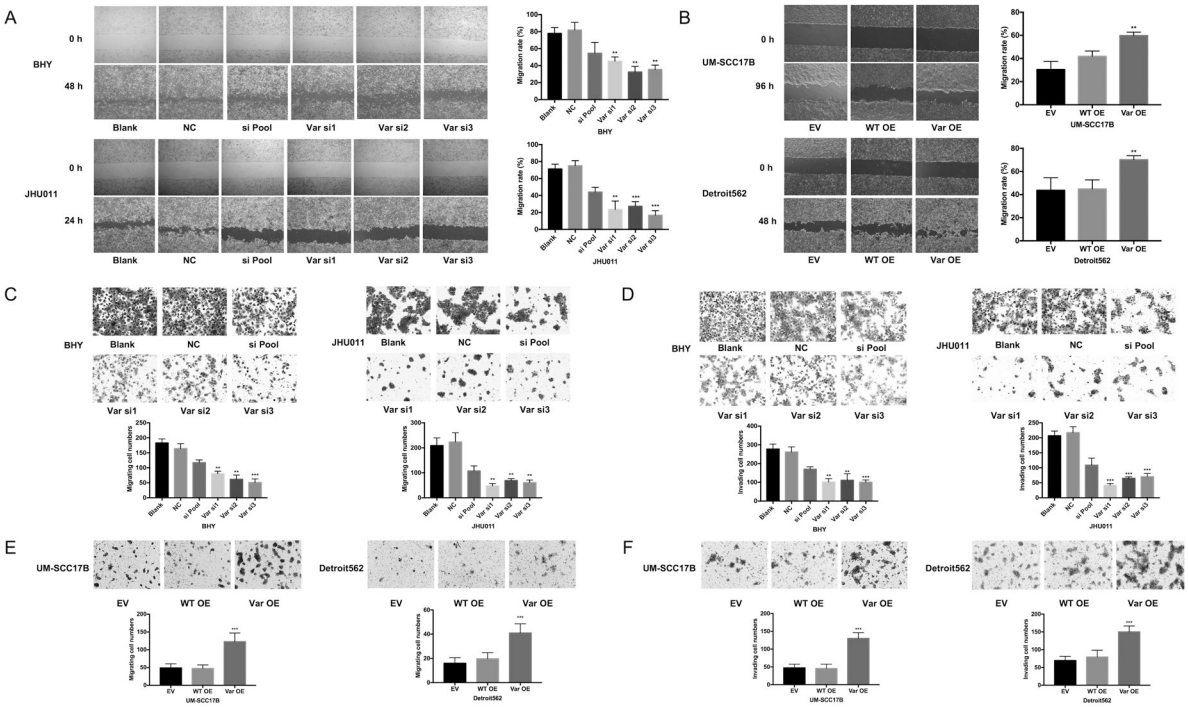

Figure 4. DOCK5 variant promotes migration and invasion of HPV negative HNSCC cells

(A) Cell scratch assay demonstrates that cell migration ability was inhibited by knockdown of DOCK5 variant in BHY and JHU011 cells. (B) Cell scratch assay demonstrates that cell migration ability was enhanced by overexpression of DOCK5 variant in UM-SCC17B and Detroit562 cells. (C) Transwell migration assay shows that the number of migrating cells is decreased by knockdown of DOCK5 variant. (D) Transwell invasion assay shows that the number of invading cells is significantly reduced by knockdown of DOCK5 variant. (E) Transwell migration assay shows that the number of migrating cells is increased by overexpression of DOCK5 variant. (F) Transwell invasion assay shows that the number of invading cells is significantly increased by overexpression of DOCK5 variant. ***: $P < 0.001$, **: $P < 0.01$.

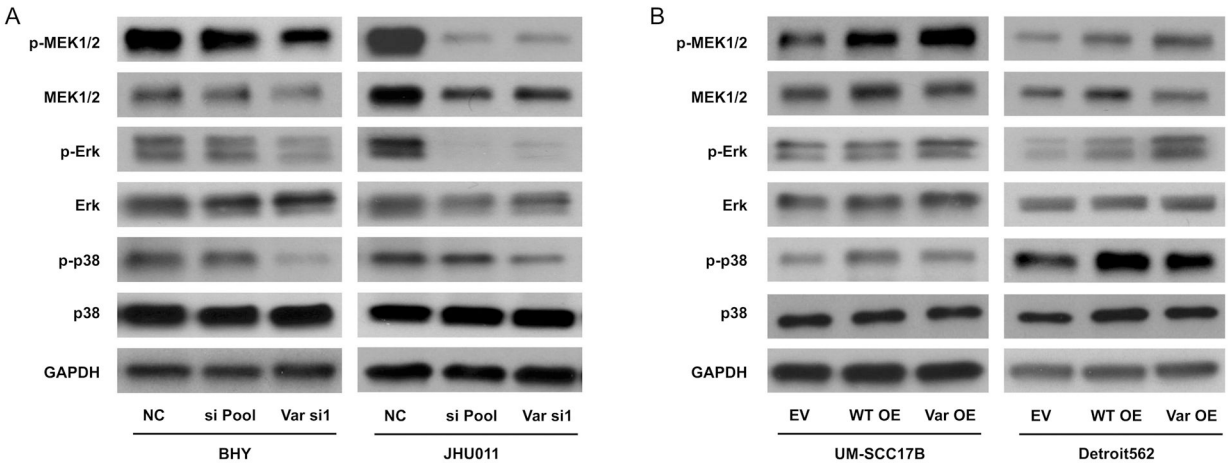

Figure 5. DOCK5 variant activates p38 and Erk MAPK pathway

(A) The expression of p-p38, p-Erk and p-MEK1/2 is decreased following knockdown the expression of DOCK5 variant in both BHY and JHU011 cells. (B) Overexpression of DOCK5 variant up-regulated the expression of p-p38, p-Erk and p-MEK1/2 in both UM-SCC17B and Detroit562 cells.

Table 1

Category of identified 580 ASEs associated with HPV negative HNSCC

Types of splicing events	No. of junction (580)	Percentage
Alternative start site	193	33.3%
Alternative end site	117	20.2%
Canonical variant or exon skip	64	11.0%
Insertion	99	17.1%
Deletion	78	13.4%
Non-coding	23	4.0%
Intron retention	6	1.0%