Journal of Pathology
J Pathol 2020; **251:** 353–364
Published online 9 July 2020 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/path.5467

ORIGINAL PAPER

Exome sequencing identifies new somatic alterations and mutation patterns of tongue squamous cell carcinoma in a Chinese population

Heyu Zhang[1,6,7†], Yongmei Song[5†], Zhenglin Du[3], Xuefen Li[1], Jianyun Zhang[4,7], Shuai Chen[4], Feng Chen[1], Tiejun Li[1,4,6,7*] and Qimin Zhan[2,5*]

[1] *Central Laboratory, Peking University School and Hospital of Stomatology, Beijing, PR China*
[2] *Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Laboratory of Molecular Oncology, Peking University Cancer Hospital & Institute, Beijing, PR China*
[3] *China National Center for Bioinformation & National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, PR China*
[4] *Department of Oral Pathology, Peking University School and Hospital of Stomatology, Beijing, PR China*
[5] *State Key Laboratory of Molecular Oncology, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, PR China*
[6] *National Clinical Research Center for Oral Diseases, Peking University School and Hospital of Stomatology, Beijing, PR China*
[7] *Research Unit of Precision Pathologic Diagnosis in Tumors of the Oral and Maxillofacial Regions, Chinese Academy of Medical Sciences (2019RU034), Beijing, PR China*

**Correspondence to: Q Zhan, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Laboratory of Molecular Oncology, Peking University Cancer Hospital & Institute, NO. 52, Beijing 100142, PR China. E-mail: zhanqimin@bjmu.edu.cn; or T Li, Department of Oral Pathology, Peking University School and Hospital of Stomatology, NO.22, Beijing 100081, PR China. E-mail: litiejun22@vip.sina.com*

[†]These authors contributed equally to this work.

Abstract

Tongue squamous cell carcinoma (TSCC) is an aggressive group of tumors characterized by high rates of regional lymph node metastasis and local recurrence. Emerging evidence has revealed genetic variations of TSCC across different geographical regions due to the impact of multiple risk factors such as chewing betel-quid. However, we know little of the mutational processes of TSCC in the Chinese population without the history of chewing betel-quid/tobacco. To explore the mutational spectrum of this disease, we performed whole-exome sequencing of sample pairs, comprising tumors and normal tissue, from 82 TSCC patients. In addition to identifying seven previously known TSCC-associated genes (*TP53*, *CDKN2A*, *PIK3CA*, *NOTCH1*, *ASXL1*, *USH2A*, and *CSMD3*), the analysis revealed six new genes (*GNAQ*, *PRG4*, *RP1*, *ZNF16*, *NBEA*, and *PTPRC*) that had not been reported previously in TSCC. Our *in vitro* experiments identified *ZNF16* for the first time as a solid tumor associated gene to promote malignancy of TSCC cells. We also identified a microRNA (miR-585-5p) encoded by the 5q35.1 region and characterized it as a tumor suppressor by targeting *SOX9*. At least one non-silent mutation of genes involved in the 10 canonical oncogenic pathways (Notch, RTK-RAS, PI3K, Wnt, Cell cycle, p53, Myc, Hippo, TGFβ, and Nrf2) was found in 82.9% of cases. Collectively, our data extend the spectrum of TSCC mutations and define novel diagnosis markers and potential clinical targets for TSCC.

Keywords: exome sequencing; tongue squamous cell carcinoma; mutation; ZNF16; miR-585-5p

Received 2 September 2019; Accepted 7 May 2020

No conflicts of interest were declared.

Introduction

Tongue squamous cell carcinoma (TSCC) is the major subtype of head and neck squamous cell carcinoma (HNSCC) and is associated with significant morbidity and relatively low overall 5-year survival rates due to its etiological and biological heterogeneity that is influenced by distinct risk factors [1,2]. Death rates of TSCC even increased slightly in men in recent years [2,3]. There are no predictive biomarkers for early diagnosis. Surgery remains the cornerstone of treatment for TSCC. Recent studies have demonstrated increasing incidence among young non-drinking and non-smoking individuals, the molecular basis of which has not been elucidated [4].

Previous genome-wide profiling studies of HNSCC or oral squamous cell carcinoma (OSCC) uncovered various genetic alterations, which have aided our

understanding of oral tumorigenesis [5–9]. However, distinct clinical behaviors and genomic variations are present between different subsites of HNSCC or OSCC [3,6]. Especially, the mutation patterns of TSCC are suggested to be different from those of other HNSCC subtypes due to the specific anatomic location and tissue characteristics [10,11]. For example, *DST* and *RNF213* mutation frequencies showed higher in TSCC than those in OSCC, whereas *TP53*, *CDKN2A*, and *NOTCH1* mutations were fewer in TSCC than in OSCC [11].

Recent exome-sequencing studies on TSCC are confined by the small sample size and therefore limited to discovery of rare gene mutations [10]. Moreover, lifestyle, diet habits, and ethnicity can also affect the genomic landscape of TSCC, and TSCC etiology varies across geographical regions [9,12]. For example, chewing betel-quid, which is known to cause oral cancer, is traditional and popular in India, and Indian TSCC exhibited different mutation patterns compared to Singapore TSCC [10,11,13]. In addition, most of the participants in the previous studies were of European descent, and human papilloma virus (HPV) infection accounted for almost 50% of cases [6]. Here, we characterize the mutational landscape of TSCC, the anatomically homogeneous cancer in Han Chinese patients who did not have the habit of chewing betel-quid/tobacco, to annotate putative cancer drivers or pathways contributing to TSCC tumorigenesis.

Materials and methods

Sample collection and patient characteristics

The ethics approval for this study was obtained from the institutional review board of the Stomatology Hospital of Peking University (Approval No. PKUSSIRB-201626006). Tumors and paired normal tissues were collected from TSCC patients who had not been treated with chemotherapy or radiotherapy before the operation during 2009–2014 in the Stomatology Hospital of Peking University. None of the patients have the history of chewing betel-quid and tobacco. Patient characteristics are summarized in supplementary material, Table S1. Recurrence and remote metastasis were important factors for prognosis (supplementary material, Table S2). In addition, 59 fresh-frozen TSCC paired tumor and normal tissues were collected for gene expression analysis later.

Exome sequencing and data analysis

We followed a previously described protocol for exome capture, library preparation, and sequencing, with some modifications (see Supplementary materials and methods) [14].

Microarrays

Use of PrimeView™ Human Gene Expression Arrays (Affymetrix) was performed by the CapitalBio Corporation (Beijing, PR China) and all primary data are available at the Gene Expression Omnibus (GEO accession: GSE114093).

Mouse xenograft assays

All animal experiments were performed with the approval and under the supervision of Peking University Biomedical Ethic Committee (Approval No. LA2017174). Two million cells of stably transfected CAL27 cells suspended in 100 µl of sterile phosphate-buffered saline (PBS) were subcutaneously injected into the right flank of 5-week-old female BALB/c nude mice. Tumor growth was monitored every 4 days using a linear caliper, and the tumor volume was calculated using the formula $V = (a \times b^2)/2$, where a is the larger dimension and b is the perpendicular diameter. The mice were sacrificed 30 days after the xenografts were seeded. Formalin-fixed paraffin-embedded xenograft tumors were stained immunohistochemically for SOX9.

Statistical analyses

The SPSS Statistics 17.0 package was employed to correlate clinical and biological variables by means of Fisher's test or a non-parametric test when necessary. A Student's *t*-test was used to analyze the results, which were expressed as means ± SD. Pearson correlation analysis was performed to evaluate the association between miR-585-5p expression and SOX9 expression and to calculate coefficient (r) and *P* values. The R package of Survival was used for survival analysis. The following indicators of statistical significance were used: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ versus control.

Additional details may be found in the Supplementary materials and methods.

Results

Overview of TSCC somatic mutation profile

To identify somatic mutations, we performed whole-exome capture on paired DNA samples from 82 TSCC patients. All samples were sequenced with a mean coverage of 125×, and more than 95% of targeted regions were covered above 30× (supplementary material, Figure S1).

We identified 5139 SNVs and 107 InDels affecting 3956 genes (supplementary material, Table S3). The average mutation rate was 1.14 per megabase (Mb) (range 0.05–3.17) and lower than that in HNSCC, esophageal squamous cell cancer (ESCC), and lung squamous cell carcinoma (LUSC) (supplementary material, Table S4) [6,14,15]. A higher number of non-synonymous mutations was observed in male patients and in smoking patients (supplementary material, Figure S2). Ninety-one of 111 (82%) of mutation sites included in 13 genes were verified by mass or Sanger sequencing (supplementary material, Table S5).

G·C >A·T transitions were the most common mutations in TSCC, followed by C·G >A·T and C·G >G·C

transversions. Compared to HNSCC, ESCC, and LUSC, the overall mutation spectrum in TSCC was most similar to that in HNSCC (supplementary material, Figure S3).

Copy number analysis from exome data identified a median of 34 CNAs per sample (range of 7–467) in TSCC (supplementary material, Tables S6 and S7 and Figure S4). Furthermore, 886 genes in 28 significantly amplified regions and 214 genes in four significantly deleted regions were uncovered (supplementary material, Tables S8 and S9). We observed recurrent gains in chromosome 3q29, 5p14.2, 6q12, 7p11.2, 8q11.1, 8q23.3, 9p24, 11q22, and 21p11.2 and recurrent deletions involving tumor suppressor genes in 9p21.3 (*CDKN2A*; 16/82), 3p14.3 (*WNT5A*; 10/80), 5q32 (*SPINK7*; 4/82), and 18q21.33 (*PHLPP1*; 9/82). Among these 32 focal CNAs, two significantly amplified regions (5p11, $p = 0.033$; 19q11, $p = 0.033$, Fisher's exact test) in TSCC were associated with regional lymph node involvement. Gain of 4p11 ($p = 0.013$, Fisher's exact test) and loss of 18q21.33 ($p = 0.045$, Fisher's exact test) were associated with regional TNM stage. Compared with HNSCC, ESCC, and LUSC, the CNA landscape of TSCC was most similar to that of HNSCC (supplementary material, Table S10 and Figure S5). We identified five CNAs (gain of 7p11.2, 11q13.3, 9p13.3, and 18q11.2; loss of 9p21.3) in all four cancers. These findings, together with mutation spectrum analysis, suggest that squamous cell carcinomas of different tissues might have common characteristics, and therapeutic approaches targeting these mutual changes could be developed.

Context-specific mutation spectrum analysis identified three distinct clusters (supplementary material, Figure S6). Cluster 1 was associated with high mutation rate ($p = 0.0058$, Fisher's exact test; supplementary material, Table S11). Cluster 3 was significantly enriched in lymph node metastasis ($p = 0.0285$, Fisher's exact test).

Virus detection in TSCC

A viral etiology has been suggested for some carcinomas [16]. We thus determined the presence of several viruses including human papilloma virus (HPV) and human herpes virus (HHV). We found that only one case arising from the base of tongue was associated with HPV infection, which is consist with previous studies reporting that the rate of HPV infection could be more frequent in tonsil subsite of HNSCC than other subsites (supplementary material, Table S12) [6]. HHV-4 (Epstein–Barr Virus, EBV), which is associated with several human malignancies as a potentially oncogenic virus, was identified in 15 cases [17]. There is inconsistent evidence regarding the involvement of EBV in TSCC pathogenesis and more research is required to conclusively establish this association [18,19].

Significantly mutated genes in TSCC

Analysis using MutSigCV and the method of Lawrence *et al* identified 13 significantly mutated genes with q values <0.1, including 7 well-known TSCC driver genes (*TP53*, *CDKN2A*, *PIK3CA*, *NOTCH1*, *ASXL1*, *USH2A*, and *CSMD3*) and 6 genes (*PRG4, ZNF16, GNAQ, NBEA, PTPRC* and *RP1*) that had not been described previously in TSCC (Figure 1 and supplementary material, Table S13).

TP53 somatic mutations were identified in 62.2% of samples. Of the 55 *TP53* mutations, 24 were truncating mutations (15 nonsense, 4 InDels, 5 splice-site) and 31 were missense. All missense mutations were located in DNA-binding domain (95–288 aa) (supplementary material, Figure S7). Twenty nine of 31 (93.5%) of *TP53* missense mutations were loss of transactivation capacity and the other 2 were partially functional as measured in yeast-based assays (supplementary material, Table S14) [20,21]. There were four recurrent mutations (p.Arg273Cys, p.Arg248Gln, p.Arg282Trp, and p.Arg342*) (n = 3), all of which were well-known cancer hotspots [22]. We found *PIK3CA*, another well-known cancer-associated gene, mutated in 7.3% of patients. Both the hotspot mutation (n = 3, c.3140A>G) and another mutation (c.3141T>A) result in the change of amino acid His1047 (supplementary material, Figure S7). Although *CSMD3*, which encodes a transmembrane protein with CUB and sushi multiple domains, was recently implicated in HNSCC [7], its function and the mechanisms are unknown. We found that *CSMD3* mutations were related to pathological grade (n = 82, $p = 0.025$, Fisher's exact test) and associated with poor prognosis in TSCC (supplementary material, Figure S8), suggesting that CSMD3 may play an important role in TSCC tumorigenesis.

PRG4 contained five missense mutations, with two of them in the same residue c.1789A>C encoding p.(Thr597Pro), probably indicating a mini-hotspot (supplementary material, Figure S7). PRG4 in different cells possesses different or even opposite functions. The recombinant human PRG4 exerted an anti-proliferative effect on rheumatoid arthritis fibroblast-like synoviocytes (RA-FLS) mediated by its interaction with CD44, resulting in a downstream inhibition of NFκB nuclear translocation [23]. However, other studies suggest that PRG4 may contribute to maintain myxoid liposarcoma cell growth through repression of IL-24 expression [24]. The function of PRG4 and its mutations have not been reported in TSCC. We first correlated the microtubule-associated protein *RP1*, the mutations in which can cause retinitis pigmentosa diseases, with tumors [25]. *RP1* was mutated in five cases and amplified in 18.3% of TSCC cases. The mutations of *PTPRC* and *NBEA* were reported in solid tumors for the first time. Both the *PTPRC* mutation c.1501C>T encoding p.(Arg501Trp) that was mapping to an extracellular domain and the *NBEA* mutation c.1910 T > C encoding p.(Val637Ala) had high PolyPhen-2 scores of 1.00 and 0.99, respectively, which predict that the variants are strongly damaging. *GNAQ*, which encodes a Gα subunit of heterotrimeric G-proteins, has been implicated as an oncogene in uveal melanoma [26]. In our studies, the *GNAQ* mutation c.286A>T encoding p.(Thr96Ser) was validated as a SNP and the mutation c.548G>A encoding p.(Arg183Gln) was found to be associated with

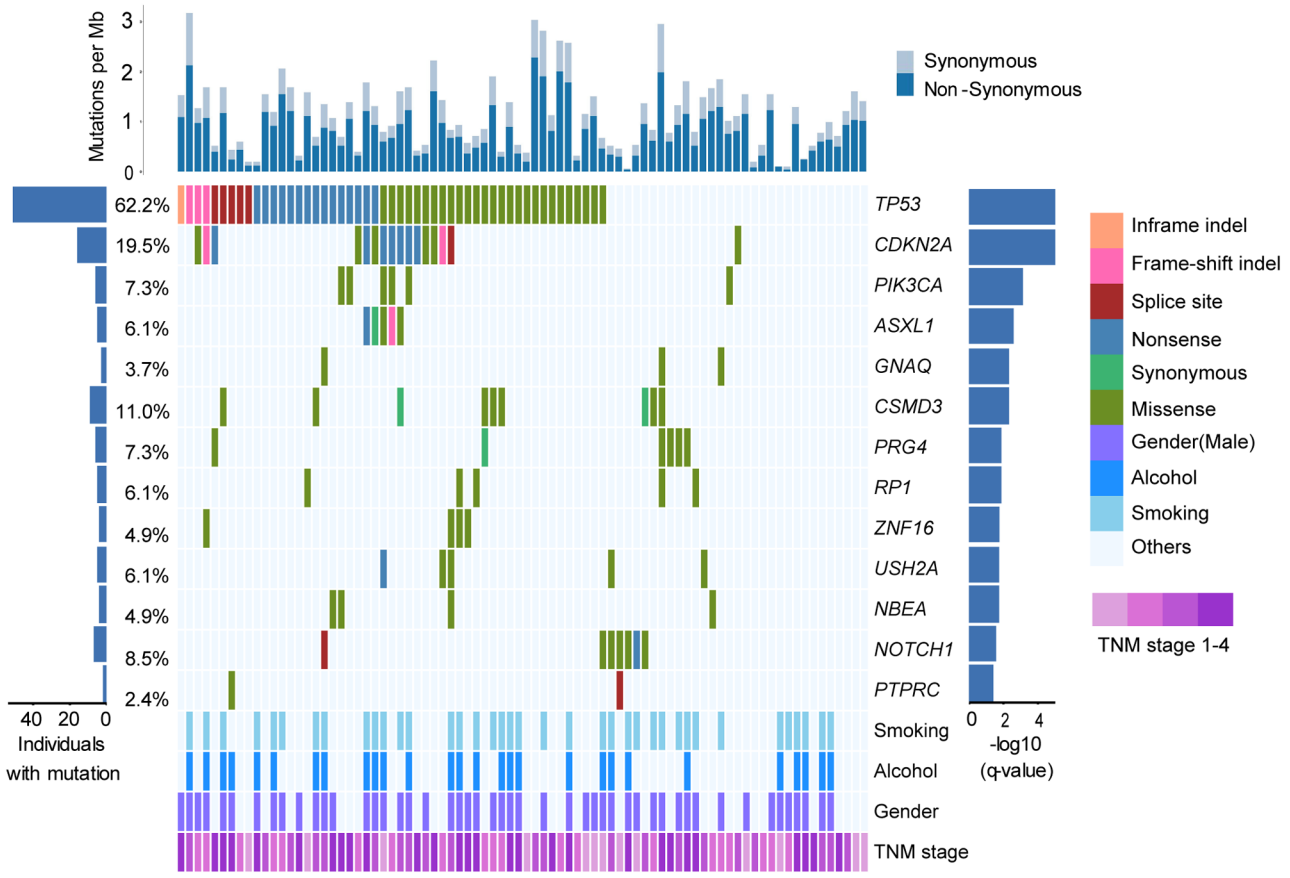

Figure 1. Significantly mutated genes in TSCC. The top panel shows the rate of synonymous and non-synonymous mutations expressed in number of mutations per megabase (Mb) of covered target sequence. The middle panel shows the matrix of mutations in significantly mutated genes colored by the type of coding mutation. Each column denotes an individual tumor, and each row represents a gene. The left panel shows the number and percentage of samples with mutations in the corresponding gene. The right panel shows significantly mutated genes listed vertically by *q*-value. The low panel shows the clinic features of TSCC patients.

port-wine stains (PWSs) and Sturge–Weber syndrome [27,28].

We studied the temporal ranking of mutations during clonal evolution using the PyClone. All mutations were evaluated, and multiple clones were found in 33 samples (median 1, range 1–7) (supplementary material, Figure S9A). Mutations belonging to the largest cluster were defined as dominant mutations, and other clusters as secondary mutations [29]. In total, 3606 (76%) and 1129 (24%) mutations were judged to be dominant and secondary, respectively (supplementary material, Table S15). Thirteen significantly mutated genes were ranked based on the amount of dominant mutations (supplementary material, Figure S9B). There was no statistically significant difference between frequencies of dominant and secondary mutations of all 13 genes (supplementary material, Figure S9C).

The above variants, which lacked functional validation, were categorized as being of unknown significance. Future studies need to validate whether these variants contribute to TSCC tumorigenesis.

ZNF16 acts as an oncogene in TSCC

We first reported the mutations of *ZNF16* (zinc finger protein 16) in malignancy. *ZNF16* was mutated in 4.9% (4 of 82) of cases. In addition, *ZNF16* was located in 8q24, which was amplified in 35.4% (29/82) of TSCCs. *ZNF16* contained four missense mutations, of which c.1354C>T encoding p.(Arg452Trp) was found in three other cancer types in the COSMIC database and is predicted to be strongly damaging (PolyPhen-2 score = 1.000). Previous studies reported that the overexpression of wild-type ZNF16 accelerated the growth rate in K562 cells [30]. However, the biological functions of ZNF16 in TSCC have been unknown.

We used an immunohistochemical method to evaluate whether the CNA amplification of *ZNF16* was related to the expression change of ZNF16 (Figure 2A). As expected, ZNF16 was upregulated in 69% (20/29) of TSCC tumor specimens harboring CNA gain of *ZNF16* than paired normal tissues. ZNF16 was highly expressed in all TSCC cell lines compared to that in control (Figure 2B). ZNF16 knockdown significantly diminished cell proliferation, migration, and invasion *in vitro* and tumor growth *in vivo* (Figure 2C–H). ZNF16 knockdown could suppress the ERK1/2 pathway (Figure 2I). We then constructed wild-type and mutant (c.1354C>T) ZNF16 plasmids and transiently transfected them into TSCC cells. Consistently, overexpression of wild-type ZNF16 significantly promoted cell

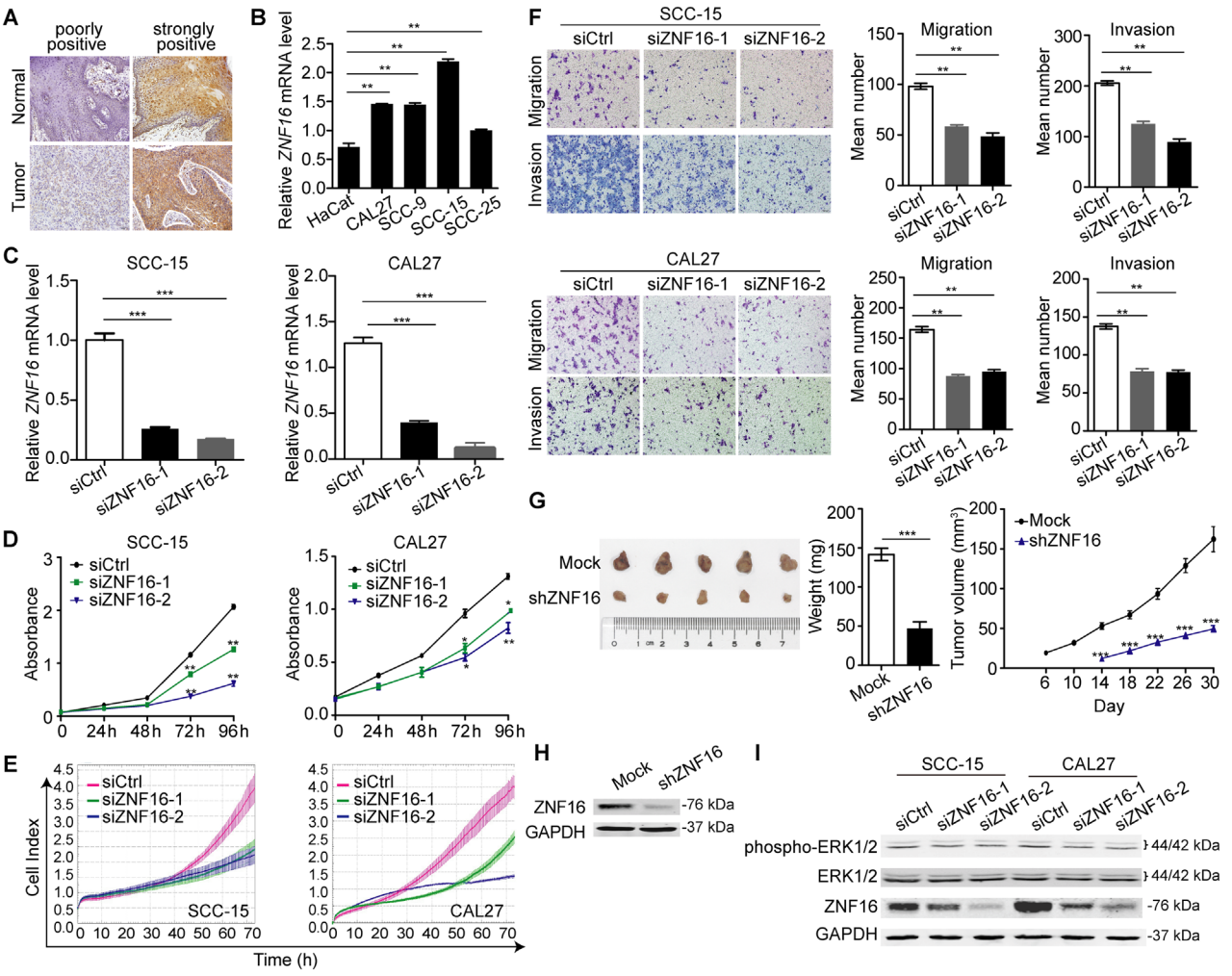

Figure 2. Knockdown of ZNF16 negatively modulates TSCC cellular malignant phenotypes. (A) Representative images of ZNF16 expression in TSCC tumor tissues and the corresponding adjacent normal tissues shown using immunohistochemistry. Scale bar: 50 μm. (B) RT-qPCR analysis of *ZNF16* expression in TSCC cell lines and an immortalized normal epithermal cell line (HaCat). (C) RT-qPCR analysis of *ZNF16* expression after siRNA knockdown in SCC-15 and CAL27 cell lines. (D) CCK-8 assay, (E) RTCA-iCELLigene system assay, and (F) Transwell assay of SCC-15 and CAL27 cell lines performed after transfection with siRNA as indicated. (G) Knockdown of ZNF16 suppressed the growth of TSCC xenografts in nude mice. Left: xenografts images were showed for each group; middle: tumor weight was showed; right: tumor volume was periodically measured for each mouse and tumor growth curves were plotted. (H) Western blot of ZNF16 in CAL27 infected with shZNF16 lentiviral particles, n = 5/group. (I) Western blot of endogenous ERK1/2 and phospho-ERK1/2 protein expression levels in SCC-15 and CAL27 cells transfected with siRNAs. GAPDH served as a loading control. All experiments were performed at least three times and data were statistically analyzed by two-sided *t*-tests. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ versus control. Error bars indicate SEM.

growth, migration, and invasion compared to control, whereas the ZNF16 mutant showed stronger capability of enhancing those phenotypes compared to wild-type ZNF16 and promoted ERK1/2 phosphorylation more strongly (supplementary material, Figure S10).

miR-585-5p acts as a tumor suppressor in TSCC

The 5q35.1 region was found to be deleted in four TSCC samples, in which miR-585-5p was identified (supplementary material, Table S7). RT-qPCR analysis showed that compared to HaCat cell line, miR-585-5p expression was markedly decreased in CAL27, SCC-9, and SCC-25 cells (Figure 3A). Overexpression of miR-585-5p expression resulted in decreased cell proliferation, migration, and invasion *in vitro* and tumor growth *in vivo*, whereas miR-585-5p inhibitors enhanced TSCC cell growth and mobility (Figure 3B–H). miR-585-5p markedly downregulated N-cadherin, vimentin, CDK2, and phosphorylated ATK, and upregulated E-cadherin (Figure 3I). We next used RT-qPCR to evaluate miR-585-5p expression in 59 pairs of TSCC tumors and normal samples. Reduced expression of miR-585-5p was detected in tumor tissues from 81.36% (48/59) of the TSCC patients ($p < 0.001$, *t*-test; Figure 3J). Analysis of clinicopathological data showed that low expression of miR-585-5p was correlated with remote metastasis (supplementary material, Table S16). We observed that high miR-585-5p was associated with longer overall survival (Figure 3K).

To further investigate the mechanisms responsible for the tumor-suppressive abilities of miR-585-5p, we identified predicted target genes by using searching TargetScan and doing the mRNA microarray of the SCC-9 cells

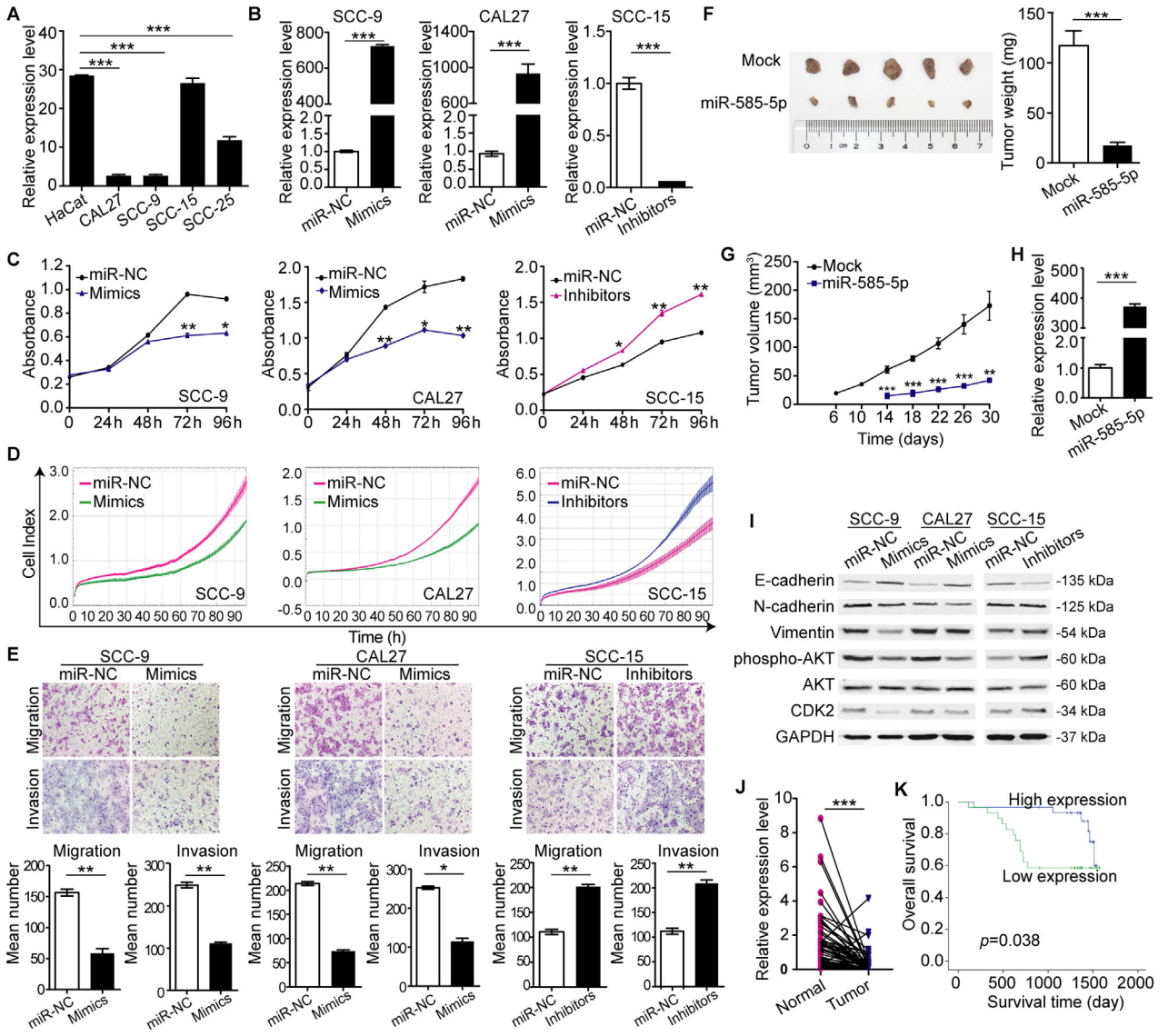

Figure 3. miR-585-5p negatively modulates TSCC cellular malignant phenotypes. (A) RT-qPCR analysis of miR585-5p in TSCC cell lines and an immortalized normal epithermal cell line (HaCat). (B) RT-qPCR analysis of miR-585-5p expression in TSCC cell lines (SCC-9, CAL27, and SCC-15) transfected with miR585-5p mimics or inhibitors as indicated. (C) CCK-8 assay, (D) RTCA-iCELLigene system assay, and (E) Transwell assay of TSCC cell lines performed after transfection with miR-585-5p mimics or inhibitors as indicated. (F) The *in vivo* effect of miR-585-5p was evaluated in xenografted mice bearing tumors originating from CAL27 cells, n = 5/group. (G) The tumor volume was measured periodically for each mouse and tumor growth curves were plotted. (H) RT-qPCR analysis of miR-585-5p expression in CAL27 infected with mock or hsa-miR-585-5p lentiviral particles. (I) Western blot of endogenous E-cadherin, N-cadherin, vimentin, phospho-AKT, AKT, and CDK2 protein expression levels in SCC-9, CAL27, and SCC-15 cells. GAPDH served as a loading control. (J) RT-qPCR analysis of miR-585-5p expression in 59 pairs of TSCC tumor tissues and the corresponding adjacent normal tissues. (K) Kaplan–Meier survival curve for miR-585-5p high-expression and miR-585-5p low-expression patients ($p = 0.038$, log-rank). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Error bars indicate SEM.

transfected with miR-585-5p mimic or control sequence (Figure 4A–D). We verified that SOX(SRY-type box)9, a high mobility group (HMG)-box transcription factor, was a potential target of miR-585-5p. Overexpression of miR-585-5p significantly inhibited luciferase activity in cells expressing SOX9-WT, whereas cells expressing SOX9-M revealed unresponsive to miR-585-5p induction (Figure 4E,F). Upregulation of miR-585-5p markedly decreased the mRNA and protein expression levels of SOX9 (Figure 4G,H). Consistent with *in vitro* results, immunohistochemistry of xenografts generated from CAL27 cells stably transfected with hsa-miR-585-5p precursor lentiviral particles revealed a marked reduction in SOX9 expression (Figure 4I). SOX9 expression was significantly upregulated in TSCC tissues compared to that in adjacent normal tissues ($p < 0.001$, *t*-test; Figure 5J), and there was a significant negative correlation between SOX9 and miR-585-5p expression in TSCC samples ($p < 0.001$, $r = -0.527$, Pearson correlation; Figure 4K). Consistent with our earlier observations, knockdown of SOX9 recapitulated the effects of the miR-585-5p mimic on TSCC cell phenotype and forced expression of SOX9 mitigated the phenotype created by mimic transfection in SCC-9 cells (supplementary material, Figure S11). These results indicate that *SOX9* was a direct target of miR-585-5p responsible for suppressing cell growth and mobility.

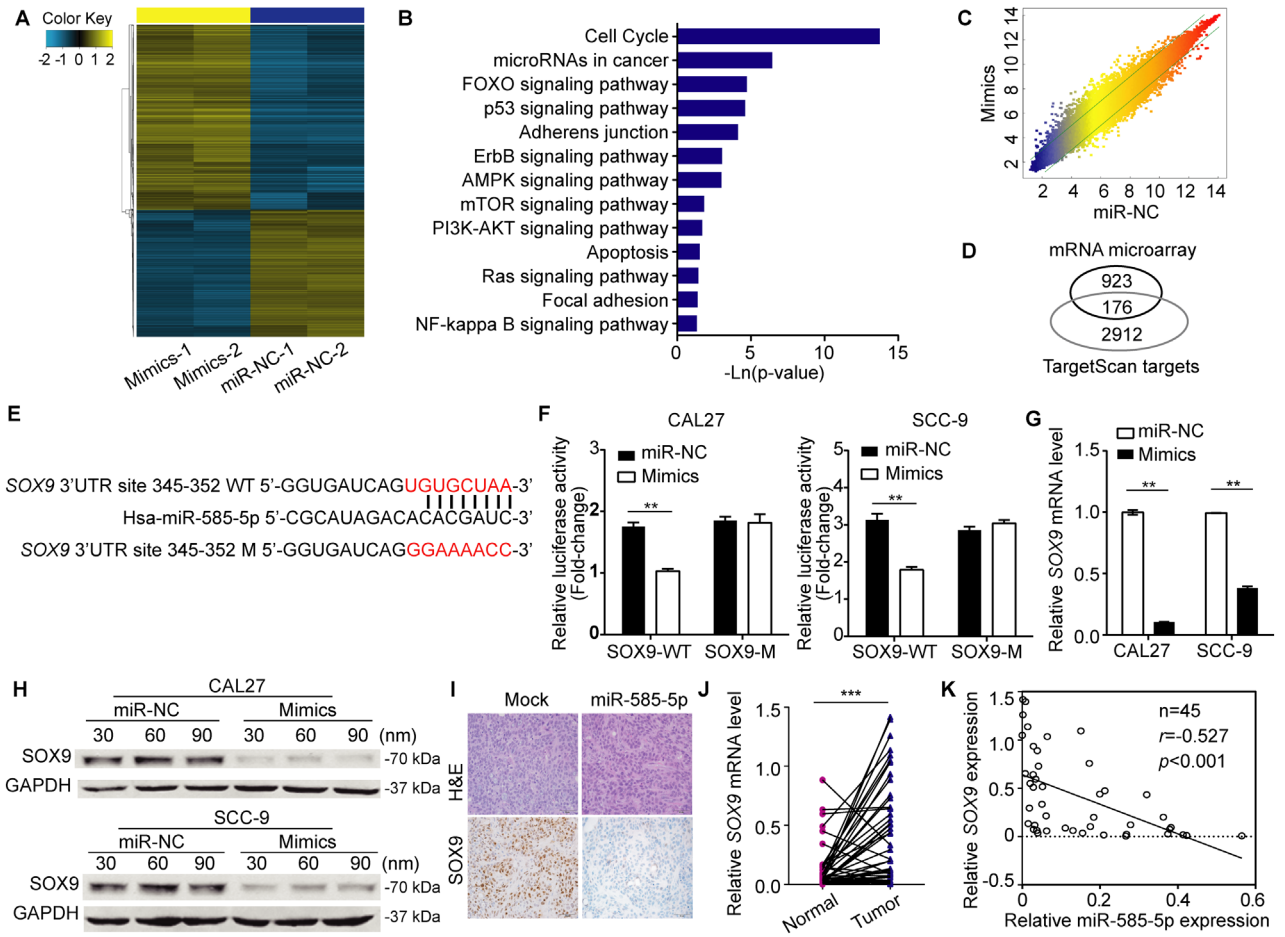

Figure 4. miR-585-5p directly regulates SOX9. (A) Cluster analysis of dysregulated genes in SCC-9 cells treated with miR-585-5p overexpression (Mimics) compared to control (miR-NC) assessed using mRNA microarrays. (B) Pathway enrichment of dysregulated genes in SCC-9 cells treated with miR-585-5p overexpression (Mimics) compared to control (miR-NC) assessed using mRNA microarrays. (C) Scatter plot presentation of genes with a cutoff of 2/0.5 in mRNA microarray analyses. (D) Venn diagrams of genes downregulated by miR-585-5p in mRNA microarray analyses and putative miR-585-5p targets predicted by TargetScan. (E) Predicted miR-585-5p target sequence in the 3′UTR of *SOX9* and the positions of mutated nucleotides. (F) Luciferase reporter assays were used to assess miR-585-5p direct targeting of *SOX9* 3′UTR in CAL27 and SCC-9 cells. (G) RT-qPCR analysis of *SOX9* expression in CAL27 and SCC-9 cells 48 h after transfected with miR-585-5p mimic or negative control. (H) Western blotting analysis of SOX9 in CAL27 and SCC-9 transfected with miR-585-5p mimic or negative control. (I) SOX9 expression was evaluated by immunohistochemistry in xenografts derived from CAL27 cells stably transfected with mock or hsa-miR-585-5p lentiviral particles. Scale bar: 50 μm. (J) RT-qPCR analysis of *SOX9* expression in 45 pairs of TSCC tissue samples. (K) Correlation analysis of *SOX9* mRNA expression and miR-585-5p expression in the paired TSCC tumor samples from 45 patients. Data shown in panels. Pearson correlation was used to calculate r and P values in panel. $^{**}p < 0.01$, $^{***}p < 0.001$. Error bars indicate SEM.

Altered pathways in TSCC

Ten canonical cancer pathways (Notch, RTK-RAS, PI3K, Wnt, Cell cycle, p53, Myc, Hippo, TGFβ, and Nrf2) are somatically altered in cancer at varying frequencies [31]. By integrating all SNVs and CNAs, we characterized the alterations of the most commonly altered genes in the 10 pathways; the alteration frequencies were detailed in Figure 5.

Of the 10 signaling pathways, the p53 pathway had the highest frequency of alterations (76.8% of cases altered). *TP53* and *ATM* accounted for 67.1% and 8.5% of cases, respectively. The genetic alterations in Hippo and RTK-RAS pathways were detected in the same frequency (70.7% of cases). The Notch pathway was altered in 65.9% of the present series, including mutations in *NOTCH1* (8.5%), *NOTCH2* (3.7%), *NOTCH3* (3.7%), and *NOTCH4* (2.4%). The mutation frequency of the Notch pathway genes in TSCC was similar to that in the squamous cell carcinomas of other sites [5,14,15]. Genetic alterations in the Wnt pathway were present in 58.5% of cases. *LRP5* and *LRP6* were mutated in 3.7% or amplified in 25.6% of cases. The mutation rate of PI3K pathway was 26.8%. *PTEN* was mutated or deleted in 6.1% of cases. The alterations in the cell cycle pathway were detected in 50% of cases. *CCND1* and *CCND2* were mutated in 2.4% and amplified in 31.7% of cases. The oxidative stress response/Nrf2 pathway was altered in 30.5% of samples and *NFE2L2* was mutated in 2.4% of cases. The TGFβ pathway had the lowest frequency of alteration (17.1% of cases). We also analyzed the enrichment of mutated genes against gene sets from the KEGG databases. Metabolic pathway ranked in the first place with 82.93% of cases accounted (supplementary material, Table S17).

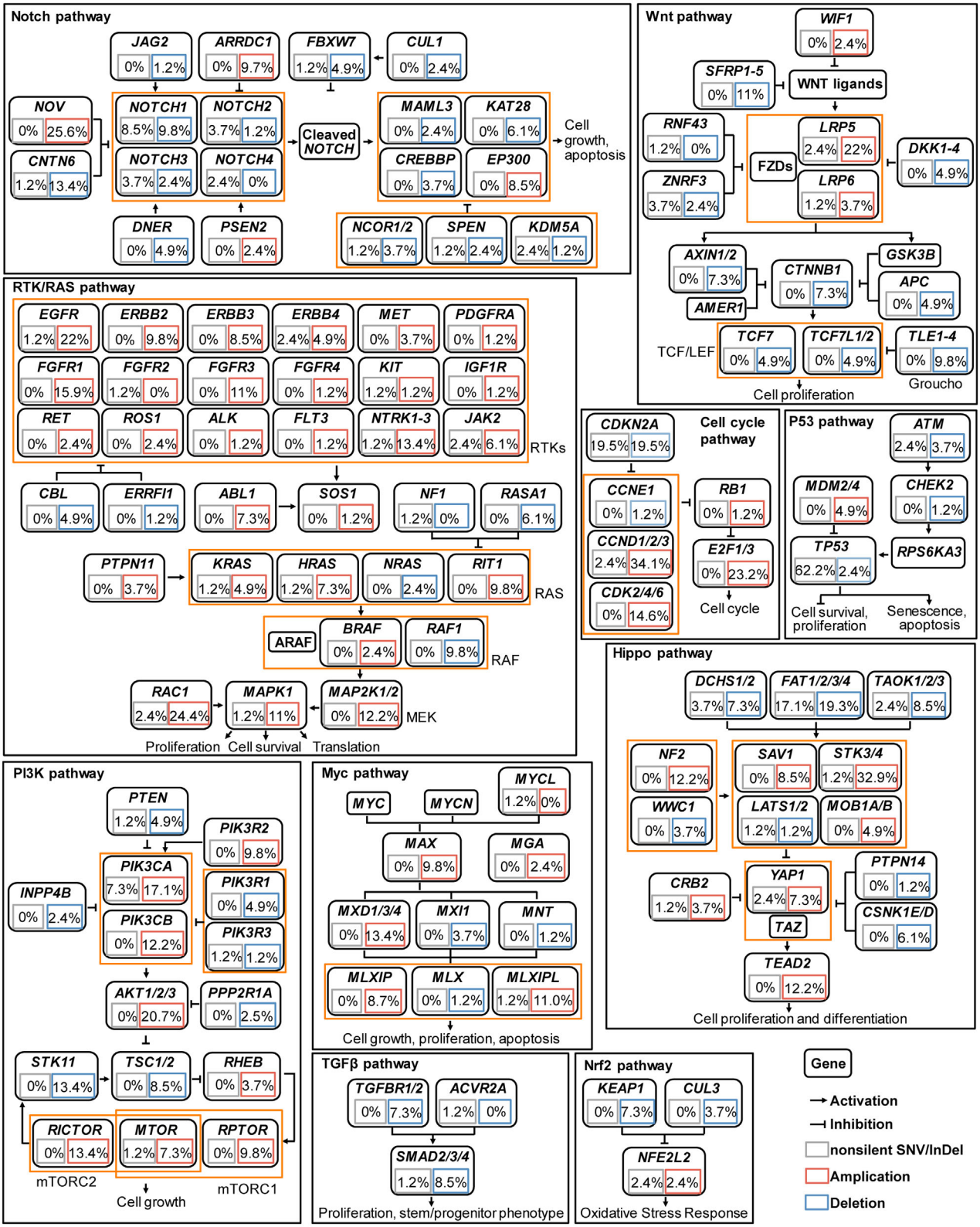

Figure 5. Somatically altered pathways in TSCC. Alteration frequencies are expressed as the percentage of cases.

Discussion

Recent large-scale sequencing studies on HNSCC included patients affected at a heterogeneous set of anatomical subsites. However, the differences in epidemiology, phenotype, and genomic alteration of HNSCC should be dissected by anatomic subsites, population, and etiology [6,32]. Our study focused on TSCC in the Chinese population, who did not have the habit of chewing tobacco or betel-quid. We identified frequent mutations in *TP53* and *CDKN2A* at similar frequencies reported in HNSCC [6,7]. However, there were fewer mutations of *PIK3CA*

(7.3% versus 21%), *FAT1* (8.5% versus 23%), *NOTCH1* (8.5% versus 19%), *KMT2D* (4.8% versus 18%), *NSD1* (1.2% versus 10%), and *CASP8* (2.4% versus 9%) in TSCC than those in HNSCC [6]. Mutation of other genes, such as *AJUBA*, *TGFBR2*, and *EPHA2* were not detected.

There are two genomic studies of TSCC reported in Indian (IN-TSCC) and Singapore (SG-TSCC) populations [10,11]. Some unique genes were detected only in SG-TSCC and not in IN-TSCC, suggesting the necessity of the parallel studies in different populations. We compared the mutational signatures of our result in which the patients were all from China (CN-TSCC) with those of Indian (IN-TSCC) and Singapore (SG-TSCC) patients (supplementary material, Figure S12). Signature 1, which has been regarded as a cell division/mitotic clock, was prevalent in all three TSCC cohorts. On a relative scale, the mutational signatures of SG-TSCC and IN-TSCC were more similar. Signature 13, which is related to APOBEC mutagenesis, was frequent in both SG-TSCC and IN-TSCC. However, signature 6, 15, and 23 were more popular in CN-TSCC. Signatures 6 and 15 are both associated with defective DNA mismatch repair and often found in microsatellite unstable tumors. Signature 23 is characterized by C > T substitutions and the etiology was unknown. Signature 29, possibly related to tobacco chewing, and signature 24, which has been associated with aflatoxin mutagens, were more prevalent in IN-TSCC. Signature 7, which is probably due to ultraviolet exposure, was also more frequent in IN-TSCC. Signature 3, which is associated with defective homologous recombination-based DNA damage repair and signature 16 (unknown etiology), were more frequent in SG-TSCC. In addition, we found that the mutation frequencies of some genes were similar between IN-TSCC and SG-TSCC, whereas they were different from CN-TSCC. For example, *TP53* and *CDKN2A* mutations are more frequently in CN-TSCC (*TP53*: 62.2%; *CDKN2A*: 19.5%) than those in IN-TSCC (*TP53*: 38%; *CDKN2A*: 6%) and SG-TSCC (*TP53*: 38%; *CDKN2A*: 5%). Moreover, *ZNF16* and *RP1* mutations were detected only in CN-TSCC. The above TSCC cohorts are all focusing on Asian populations. The mutational landscape of SG-TSCC is more similar to that of IN-TSCC, which may be because both of their patients are from south Asia and have the nearby environmental features, whereas our TSCC patients are enrolled from North China, which has the disparate environment, lifestyle, and diet habit, such as hot food and chewing betel nut.

Smoking and old age are epidemiologic risk factors for TSCC. In our study, smokers had more non-silent mutations. Our analysis did not find specific mutation associated with younger patients or non-smokers, which is consistent with previous studies [33]. Previous studies reported that young non-smokers of TSCC patients were genomically similar to older smokers of TSCC patients and suggest that the smoking signature in HNSCC was largely driven by the laryngeal subsite. These findings suggest that tobacco, despite being a tumor initiator in lung cancer, may act as a tumor promoter in TSCC. The causes for increasing incidence of TSCC in young patients are still unknown, and an understanding of the molecular mechanisms contributing to the tumorigenesis of young TSCC patients is necessary for us to explore targeted preventive and treatment measures.

Consistent with previous HNSCC or TSCC reports, *TP53* was the most frequently mutated gene in our cohort [6,11]. Studies have showed that *TP53* missense mutations were scattered throughout the coding sequence, but 97% of them clustered in exons encoding the DNA-binding domain [34]. All missense mutations in our study were located in this region and the hotspot mutations were reported to have oncogenic activities (gain of function) in addition to loss of transactivation [21]. In addition, *TP53*-mutated cases harbored significantly more somatic mutations than non–*TP53*-mutated cases, suggesting that this mutation could divided TSCC cases into two types with different *TP53* mutation backgrounds ($p = 0.001$, *t*-test).

Ten oncogenic signaling pathways were curated by The Cancer Genome Atlas (TCGA) [31]. In our sample, 82.9% of cases had at least one non-silent mutation in these pathways, and 63.4% of cases had non-silent mutations of genes involved in the p53 pathway. The PI3K and Hippo pathways were ranked with the second highest rate. Notch pathway mutations were found in 23.2% of cases. Considering both non-silent mutations and CNAs, 95.1% of cases have at least one alteration in these pathways. Over 50% of TSCC samples had the alterations in the p53, RTK-RAS, Hippo, Notch, Wnt, and PI3K pathways. Despite the incremental investigation of oncogenic roles of these pathways, the number of approved pathway-targeting drugs remains sparse. More efforts should be made to develop new effective therapies for TSCC.

Recent comprehensive analysis showed that squamous cancer types contained higher proportions of mutations of the chromatin histone modification genes [26]. We also detected frequent non-silent mutations in 60 histone modification-related genes in TSCC (supplementary material, Table S18). The most frequent alterations were observed in histone-H3–modifying lysine methyltransferases. We found most of mutated histone modification-related genes in TSCC also showed mutations in ESCC, HNSCC, and LUSC. The frequencies of mutations in TSCC are more similar with those in ESCC and lower in LUSC and HSNCC. For example, H3K27me3-related genes were altered in 58.5% of TSCC, 55.7%, of ESCC, 96.6% of LUSC, and 78.9% of HNSCC. H3K79me2-related genes were mutated in 17.1% of TSCC, 15.9% of ESCC, 58.4% of LUSC and 35.5% of HNSCC.

Based on the analysis of potential therapeutic applications, all FDA-approved cancer targets had very low frequencies of mutations in TSCC samples (supplementary material, Table S19) [35,36]. *KRAS* was non-silent mutated in 1.2% of cases. The mutations of *BRCA1*, *BRAF*, *PDGFRA*, *TSC1*, and *KIT* were not detected in our samples. In addition, the Pathways in Cancer (hsa05200) includes the greatest number of potential therapeutic target genes mutated in TSCC followed by

Focal Adhesion pathway (supplementary material, Table S20).

In this study, we have described in detail the landscape of genetic alterations in TSCC and provided new insights into the biological process underlying its tumorigenesis. The novel mutated genes identified in our study warrant further investigation to determine their biologic role in cancer formation and progression. Despite the previous genomic research in HNSCC, no new therapeutic targets have been identified. To our knowledge, this is the first study identifying targetable alterations of TSCC in a Chinese population without the habit of betel-quid chewing. These findings provide support for the development of personalized therapies targeting key oncogenically activated signaling pathways for the treatment of TSCC.

Acknowledgements

This work was supported by research grants from the National Nature Science Foundation of China (81490753, 81671006, 81030018, and 81300894), the National 973 Program (2015CB553904), the Beijing Nature Science Foundation (7172238), the Fund for Fostering Young Scholars of Peking University Health Science Center (BMU2018PY026), and CAMS Innovation Fund for Medical Sciences (2019-12M-5-038).

Author contributions statement

QZ and TL directed the study and supervised the research. HZ, YS, and ZD performed the sequencing and analyzed the data. HZ, JZ, and SC performed functional assays. XL and FC collected tumor specimens and analyzed the clinical data. QZ, TL, HZ, YS, and ZD wrote the manuscript.

Data availability statement

All microarray data are available at the Gene Expression Omnibus (GEO accession: GSE114093). The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive in BIG Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession numbers PRJCA002133 that is publicly accessible at https://bigd.big.ac.cn/gsa

References

1. Chen W, Zheng R, Baade PD, *et al*. Cancer statistics in China, 2015. *CA Cancer J Clin* 2016; **66:** 115–132.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* 2019; **69:** 7–34.
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin* 2018; **68:** 7–30.
4. Annertz K, Anderson H, Biorklund A, *et al*. Incidence and survival of squamous cell carcinoma of the tongue in Scandinavia, with special reference to young adults. *Int J Cancer* 2002; **101:** 95–99.
5. Agrawal N, Frederick MJ, Pickering CR, *et al*. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* 2011; **333:** 1154–1157.
6. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* 2015; **517:** 576–582.
7. Stransky N, Egloff AM, Tward AD, *et al*. The mutational landscape of head and neck squamous cell carcinoma. *Science* 2011; **333:** 1157–1160.
8. Morris LG, Chandramohan R, West L, *et al*. The molecular landscape of recurrent and metastatic head and neck cancers: insights from a precision oncology sequencing platform. *JAMA Oncol* 2017; **3:** 244–255.
9. Su SC, Lin CW, Liu YF, *et al*. Exome sequencing of Oral squamous cell carcinoma reveals molecular subgroups and novel therapeutic opportunities. *Theranostics* 2017; **7:** 1088–1099.
10. Krishnan N, Gupta S, Palve V, *et al*. Integrated analysis of oral tongue squamous cell carcinoma identifies key variants and pathways linked to risk habits, HPV, clinical parameters and tumor recurrence. *F1000Res* 2015; **4:** 1215.
11. Vettore AL, Ramnarayanan K, Poore G, *et al*. Mutational landscapes of tongue carcinoma reveal recurrent mutations in genes of therapeutic and prognostic relevance. *Genome Med* 2015; **7:** 98.
12. Alexandrov LB, Ju YS, Haase K, *et al*. Mutational signatures associated with tobacco smoking in human cancer. *Science* 2016; **354:** 618–622.
13. India Project Team of the International Cancer Genome C. Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. *Nat Commun* 2013; **4:** 2873.
14. Song Y, Li L, Ou Y, *et al*. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature* 2014; **509:** 91–95.
15. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012; **489:** 519–525.
16. Mesri EA, Feitelson MA, Munger K. Human viral oncogenesis: a cancer hallmarks analysis. *Cell Host Microbe* 2014; **15:** 266–282.
17. Tsao SW, Tsang CM, To KF, *et al*. The role of Epstein–Barr virus in epithelial malignancies. *J Pathol* 2015; **235:** 323–333.
18. Jiang R, Gu X, Moore-Medlin TN, *et al*. Oral dysplasia and squamous cell carcinoma: correlation between increased expression of CD21, Epstein-Barr virus and CK19. *Oral Oncol* 2012; **48:** 836–841.
19. Wilms T, Khan G, Coates PJ, *et al*. No evidence for the presence of Epstein-Barr virus in squamous cell carcinoma of the mobile tongue. *PLoS One* 2017; **12:** e0184201.
20. Kato S, Han SY, Liu W, *et al*. Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc Natl Acad Sci U S A* 2003; **100:** 8424–8429.
21. Petitjean A, Mathe E, Kato S, *et al*. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum Mutat* 2007; **28:** 622–629.
22. Bouaoun L, Sonkin D, Ardin M, *et al*. TP53 variations in human cancers: new lessons from the IARC TP53 database and genomics data. *Hum Mutat* 2016; **37:** 865–876.
23. Al-Sharif A, Jamal M, Zhang LX, *et al*. Lubricin/proteoglycan 4 binding to CD44 receptor: a mechanism of the suppression of proinflammatory cytokine-induced synoviocyte proliferation by lubricin. *Arthritis Rheumatol* 2016; **67:** 1503–1513.
24. Oikawa K, Mizusaki A, Takanashi M, *et al*. PRG4 expression in myxoid liposarcoma maintains tumor cell growth through suppression of an antitumor cytokine IL-24. *Biochem Biophys Res Commun* 2017; **485:** 209–214.

25. Al-Rashed M, Abu Safieh L, Alkuraya H, *et al*. RP1 and retinitis pigmentosa: report of novel mutations and insight into mutational mechanism. *Br J Ophthalmol* 2012; **96:** 1018–1022.
26. Bailey MH, Tokheim C, Porta-Pardo E, *et al*. Comprehensive characterization of cancer driver genes and mutations. *Cell* 2018; **173:** 371–385.e318.
27. Lian CG, Sholl LM, Zakka LR, *et al*. Novel genetic mutations in a sporadic port-wine stain. *JAMA Dermatol* 2014; **150:** 1336–1340.
28. Shirley MD, Tang H, Gallione CJ, *et al*. Sturge–Weber syndrome and port-wine stains caused by somatic mutation in GNAQ. *N Engl J Med* 2013; **368:** 1971–1979.
29. Nagata Y, Makishima H, Kerr CM, *et al*. Invariant patterns of clonal succession determine specific clinical features of myelodysplastic syndromes. *Nat Commun* 2019; **10:** 5m386.
30. Li XB, Chen J, Deng MJ, *et al*. Zinc finger protein HZF1 promotes K562 cell proliferation by interacting with and inhibiting INCA1. *Mol Med Rep* 2011; **4:** 1131–1137.
31. Sanchez-Vega F, Mina M, Armenia J, *et al*. Oncogenic signaling pathways in The Cancer Genome Atlas. *Cell* 2018; **173:** 321–337.e310.
32. Bundgaard T, Bentzen SM, Sogaard H. Histological differentiation of oral squamous cell cancer in relation to tobacco smoking. *Eur J Cancer B Oral Oncol* 1995; **31B:** 118–121.
33. Pickering CR, Zhang J, Neskey DM, *et al*. Squamous cell carcinoma of the oral tongue in young non-smokers is genomically similar to tumors in older smokers. *Clin Cancer Res* 2014; **20:** 3842–3848.
34. Olivier M, Eeles R, Hollstein M, *et al*. The IARC TP53 database: new online mutation analysis and recommendations to users. *Hum Mutat* 2002; **19:** 607–614.
35. Li MM, Datto M, Duncavage EJ, *et al*. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J Mol Diagn* 2017; **19:** 4–23.
36. Chakravarty D, Gao J, Phillips SM, *et al*. OncoKB: a precision oncology Knowledge Base. *JCO Precis Oncol* 2017; **2017**.
37. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009; **25:** 1754–1760.
38. McKenna A, Hanna M, Banks E, *et al*. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20:** 1297–1303.
39. Cibulskis K, Lawrence MS, Carter SL, *et al*. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013; **31:** 213–219.
40. Saunders CT, Wong WS, Swamy S, *et al*. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 2012; **28:** 1811–1817.
41. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38:** e164.
42. Huang D-W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; **4:** 44–57.
43. Lawrence MS, Stojanov P, Polak P, *et al*. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 2013; **499:** 214–218.
44. Kan Z, Jaiswal BS, Stinson J, *et al*. Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* 2010; **466:** 869–873.
45. Chiang DY, Getz G, Jaffe DB, *et al*. High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat Methods* 2009; **6:** 99–103.
46. Mermel CH, Schumacher SE, Hill B, *et al*. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011; **12:** R41.
47. Gao J, Aksoy BA, Dogrusoz U, *et al*. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013; **6:** pl1.
48. Roth A, Khattra J, Yap D, *et al*. PyClone: statistical inference of clonal population structure in cancer. *Nat Methods* 2014; **11:** 396–398.
49. Chen Y, Yao H, Thompson EJ, *et al*. VirusSeq: software to identify viruses and their integration sites using next-generation sequencing of human cancer tissue. *Bioinformatics* 2013; **29:** 266–267.
50. Zhang Y, Lv J, Liu H, *et al*. HHMD: the human histone modification database. *Nucleic Acids Res* 2010; **38:** D149–D154.
51. Li YH, Yu CY, Li XX, *et al*. Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res* 2018; **46:** D1121–D1127.
52. Members BIGDC. Database resources of the BIG data Center in 2019. *Nucleic Acids Res* 2019; **47:** D8–D14.
53. Wang Y, Song F, Zhu J, *et al*. GSA: genome sequence archive. *Genom Proteom Bioinf* 2017; **15:** 14–18.
54. Zhang H, Ma X, Shi T, *et al*. NSA2, a novel nucleolus protein regulates cell proliferation and cell cycle. *Biochem Biophys Res Commun* 2009; **391:** 651–658.

References 37–54 are cited only in the supplementary material.

SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods

Figure S1. Sequencing depth and coverage in 82 pairs of TSCC

Figure S2. Number of non-synonymous mutations in samples

Figure S3. Spectrum of somatic point mutations in TSCC, HNSCC, ESCC, and LUSC

Figure S4. CNAs landscape and GISTIC analysis of CNAs in TSCC

Figure S5. Comparison of CNAs in TSCC, HNSCC, LUSC, and ESCC

Figure S6. Mutation spectrum analysis of TSCC

Figure S7. Distribution of somatic mutations of significantly mutated genes

Figure S8. The relationship between survival time and mutations of *CSMD3* in TSCC patients

Figure S9. Clonal analysis of dominant and secondary mutations

Figure S10. ZNF16 positively modulates TSCC cellular malignant phenotypes

Figure S11. SOX9 is a target of miR-585-5p