scientific reports

Check for updates

OPEN

Revealing molecular mechanisms of early-onset tongue cancer by spatial transcriptomics

Marina R. Patysheva[1,14], Elena S. Kolegova[1,14]✉, Anna A. Khozyainova[1], Elizaveta A. Prostakishina[1], Vyacheslav Y. Korobeynikov[1], Maxim E. Menyailo[1,2], Pavel S. Iamshchikov[1], Dmitry M. Loos[3,4], Oleg I. Kovalev[3,4], Marina V. Zavyalova[3,4], Irina K. Fedorova[5], Denis E. Kulbakin[5], Irina V. Larionova[1,6], Andrey P. Polyakov[7], Liliya P. Yakovleva[8], Mikhail A. Kropotov[9], Natalya S. Sukortseva[10], Yusheng Lu[11], Lee Jia[12], Rohit Arora[13], Evgeny L. Choinzonov[5], Pinaki Bose[13], The Consortium Etiology and Pathogenesis of Oral Cancer in Young Adults & Evgeny V. Denisov[1,2]

Tongue cancer at a young age demonstrates an increase in incidence, aggressiveness, and poor response to therapy. Classic etiological factors for head and neck tumors such as tobacco, alcohol, and human papillomavirus are not related to early-onset tongue cancer. Mechanisms of development and progression of this cancer remain unclear. In this study, we performed spatial whole-transcriptome profiling of tongue cancer in young adults compared with older patients. Nine patients with tongue squamous cell carcinoma (T2-3N0-1M0) were included and divided into two groups: younger (n = 5) and older than 45 years (n = 4). Formalin-fixed paraffin-embedded (FFPE) and fresh frozen (FF) samples of tumor tissue from 4 young and 5 older patients, respectively, were used for spatial transcriptomic profiling using the 10 × Genomics Visium. The findings were validated using SeekGene single cell full-length RNA sequencing (1 young vs 1 older patient) and TCGA data (15 young vs 70 older patients). As a result, we performed the first successful integration of spatial transcriptomics data from FF and FFPE samples and revealed distinctive features of tongue cancer in young adults. Oxidative stress, vascular mimicry, and MAPK and JAK-STAT pathways were enriched in early-onset tongue cancer. Tumor microenvironment demonstrated increased gene signatures corresponding to myeloid-derived suppressor cells, tumor-associated macrophages, and plasma cells. The invasive front was accompanied by vascular mimicry with arrangement of tumor-associated macrophages and aggregations of plasma cells and lymphocytes organized into tertiary lymphoid structures. Taken together, these results indicate that early-onset tongue cancer has distinct transcriptomic features and molecular mechanisms compared to older patients.

Keywords Oral cancer, Tongue squamous cell carcinoma, Young adults, Spatial transcriptomics, Single cell sequencing

Abbreviations

ALRA	Adaptively-thresholded low rank approximation
AUC	Area under the curve
CAFs	Cancer-associated fibroblasts
CNA	Copy number alterations
HPV	Human papillomavirus
MAPK	Mitogen activated kinase
MDSCs	Myeloid-derived suppressor cells
LTA	Lymphotoxin alpha
PLAU	Plasminogen activator, urokinase
ROC	Receiver operating characteristic
TAMs	Tumor-associated macrophages
TCGA	The Cancer Genome Atlas
TGFb	Tumor growth factor beta
TLSs	Tertiary lymphoid structures
TSCC	Tongue squamous cell carcinoma

FF Fresh frozen
FFPE Formalin-fixed, paraffin-embedded

[1]Laboratory of Cancer Progression Biology, Cancer Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences, Tomsk, Russia. [2]Laboratory of Single Cell Biology, Research Institute of Molecular and Cellular Medicine, Peoples' Friendship University of Russia (RUDN University), Moscow, Russia. [3]Department of Pathological Anatomy, Siberian State Medical University, Tomsk, Russia. [4]Department of General and Molecular Pathology, Cancer Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences, Tomsk, Russia. [5]Department of Head and Neck Tumors, Cancer Research Institute, Tomsk National Research Medical Center, Russian Academy of Sciences, Tomsk, Russia. [6]Laboratory of Molecular Cancer Therapy, Cancer Research Institute, Tomsk National Research Medical Center Tomsk, Russian Academy of Sciences, Tomsk, Russia. [7]Microsurgery Department, P.A. Herzen Moscow Oncology Research Institute – a branch of the National Medical Research Radiological Center, Moscow, Russia. [8]Department of Head and Neck Tumors, A.S. Loginov Moscow Clinical Scientific Center, Moscow Healthcare Department, Moscow, Russia. [9]Surgical Department N10 of Head and Neck Tumors, N.N. Blokhin National Medical Research Center of Oncology, Ministry of Health of the Russian Federation, Moscow, Russia. [10]Department of Oncology, Radiotherapy and Plastic Surgery, I.M. Sechenov First Moscow State Medical University, Ministry of Health of the Russian Federation, Moscow, Russia. [11]College of Materials and Chemical Engineering, Minjiang University, Fuzhou, Fujian, China. [12]The First Affiliated Hospital, Henan University, Kaifeng, Henan, China. [13]Department of Biochemistry & Molecular Biology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada. [14]These authors contributed equally: Marina R. Patysheva, and Elena S. Kolegova. ✉email: elenakolegowa@mail.ru

Squamous cell carcinoma of the oral cavity has historically been considered a disease more common in individuals in their sixth decade of life that is related to cumulative exposure to tobacco, alcohol, and human papillomavirus (HPV) infection[1]. However, the first cases of this cancer in young adults were described in the 1930s, and the incidence of early-onset tongue squamous cell carcinoma (TSCC), especially in women, is increasing every year[2]. The etiological factors and mechanisms of TSCC initiation and progression in young adults have been the subject of significant debate and controversy owing to a lack of association with tobacco, alcohol consumption, and HPV[3].

Single studies showed that some cases with early-onset TSCC are related to the KIR2DL1$^+$-HLA-C2$^+$ genotype and MMP-1 2G allele. Molecular profile of early-onset TSCC are characterized by alterations in the *TP53, CDKN2A, CASP8, NOTCH1, FAT, ATXN1,* and *CDC42EP1* genes and methylation changes in the *RASSF1A, RASSF2A, MGMT, DAPK,* and *FHIT* genes[3–7]. Bulk RNA sequencing demonstrated upregulation of the immunosuppressive *LAG-3* and *HAVCR2* genes in early-onset TSCC[8]. By analyzing The Cancer Genome Atlas (TCGA) data, we found that TSCC in young adults displays reduced mutational burden, overexpression of the Rap1, PI3K-Akt, MAPK, JAK-STAT, cGMP-PKG and Fc-gamma R-mediated phagocytosis signaling genes, and specific microbiome[9]. Nevertheless, a comprehensive picture of the pathogenesis of TSCC in young adults is still unclear. This can be due to a lack of investigations of tumor tissue as a complex system considering both tumor and microenvironment cell populations and their interactions.

Spatial transcriptomics technologies make it possible to analyze the transcriptome of individual cells with their original localization within the tissue. A recent study developed a spatial transcriptome map of early-stage oral squamous cell carcinoma, adjacent precancerous lesions, and a matched normal region[10]. Spatial transcriptomics also revealed differences in the biology of core and edge regions of oral tumors and identified patterns of expressed genes associated with disease prognosis[11].

In this work, we set out to investigate transcriptomics features of early-onset TSCC relative to this cancer in older adults. We uncovered a set of MAPK and JAK-STAT signaling pathways genes in the tumor cells and a decrease in antigen presentation, alongside myeloid derived suppression cells and tumor-associated macrophages with their synergistic efforts promoting vascular mimicry. Plasma cells and lymphocytes in tertiary lymphoid structures (TLSs) were enriched in the invasive front of TSCC in young adults. Together, our results highlight potential mechanisms of early-onset TSCC and serve as a reference for further studies of this cancer.

Results

Spatial transcriptome profiling of early-onset TSCC

Tumor samples from 5 young and 4 older patients were used to investigate how early-onset TSCC differs from late-onset ones in spatial transcriptomics (Fig. 1A). First at all, we performed the integration of spatial transcriptomics data from FFPE and FF samples. It is worth noting that the data included in the study differed significantly in the method of transcript capture. Molecular profiling of FF samples is carried out based on total mRNA, while profiling of FFPE samples is performed using probes targeting protein-coding genes. To solve this problem, we removed genes from FF samples that were not captured by FFPE probes prior to integration. Normalization was first applied to the dataset to eliminate expression differences caused by unavoidable technical reasons such as unequal amplification during PCR. Batch effect correction was performed by Harmony that is an approved package by 10×Genomics and also named the preferred package by independent benchmark and was based on several covariates including sample type (FFPE or FF)[12] (Fig. 1A). Genes from FF samples that are not captured by FFPE probes were excluded prior to integration. Overall, such an approach allowed to perform successful integration of the FFPE and FF spatial transcriptomics data as seen in the UMAP graph (Fig. 1A) and to conduct the downstream analysis.

A total of 9 cell clusters were partitioned into seven major cell types, including two tumor cell clusters (#1 and 2), cancer-associated fibroblasts (CAFs; #3), muscle (#4), normal epithelium (#8), salivary gland (#9), lymphoid and myeloid immune cells (#5 and 6, respectively; Fig. 1B). Tumor cell clusters were distinguished based on

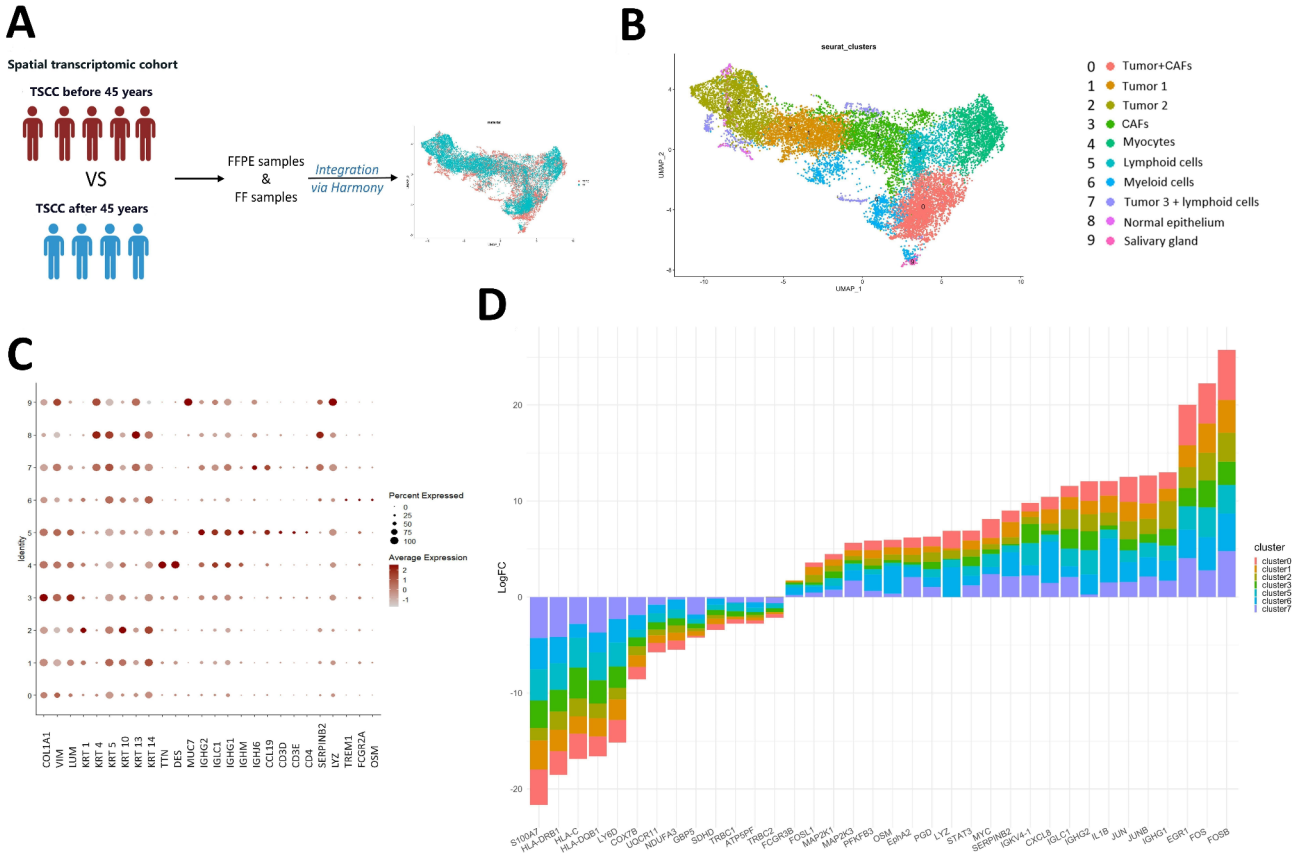

Fig. 1. Spatial transcriptomics of early-onset TSCC. A, Sampling strategy, cohort overview, and sample integration; B, Overview of the identified cell types; C, Cluster identification by cell type markers; D, Consensus plot displaying the cumulative average logFC for genes (adj. $p < 0.05$) differentially expressed between early- and late-onset TSCC: logFC > 0, genes upregulated in early-onset TSCC; logFC < 0, genes downregulated in early-onset TSCC (adj. $p < 0.05$, two-sided Wilcoxon rank-sum test, Bonferroni correction). Normal epithelium and normal salivary gland cell clusters are not shown. TSCC, tongue squamous cell carcinoma; FFPE, formalin-fixed paraffin-embedded; FF, fresh frozen; CAFs, cancer-associated fibroblasts.

the expression of the *KRT1, KRT4, KRT5, KRT10, KRT13*, and *KRT14* genes (Fig. 1C), CAFs – the COL1A1, *VIM*, and *LUM* genes, lymphoid – the *IGHG1, IGHG2, IGLC1, IGHM, IGHJ6, CCL19, CD3D, CD3E*, and *CD4* genes, and myeloid immune cells—the *SERPINE, LYZ, TREM1, FCGR2A*, and *OSM* genes (Fig. 1B,C). Two clusters were mixed with fibroblasts (#0: tumor + CAFs) and lymphoid cells (#7: tumor 3 + lymphoid cells) and demonstrated expression of the *COL1A1, VIM, KRT5, KRT14* and *KRT1, KRT4, KRT5, KRT10, KRT13, KRT14, IGHG1, IGHG2, IGLC1, IGHM, IGHJ6* genes, respectively (Fig. 1B,C).

Transcriptome features of tumor cells in early-onset TSCC

We focused on finding common features for tumor cell clusters in early-onset TSCC compared to TSCC in older patients. All tumor clusters (#0, 1, 2, and 7) were reliably distinguished by increased expression of mitogen activated kinase (MAPK) pathway genes: *EGR1, IL1B, MAP2K1, MAP2K3, FOS, FOSB, JUN, JUNB, FOSL1*, and *FOSL2* (Fig. 1D). In contrast, the expression of oxidative phosphorylation genes (*ATF5PF, SDHD, GBP5, NDUFA3, UQCR11, COX7B*, and *S100A7*) was decreased, while the *PFKFB3* and *PGD* genes, regulating glycolysis, the pentose-phosphate pathway, and glutathione metabolism, were upregulated in TSCC of young adults (Fig. 1D).

Transcriptome features of immune cells in early-onset TSCC

In comparison with late-onset TSCC, myeloid cells (#6) of tongue tumors in young adults were distinguished by increased expression of the *SERPINB2, LYZ*, and *OSM* genes corresponding to tumor-associated macrophages (TAMs) and the *FCGRB3* and *PFKFB3* genes being markers of myeloid-derived suppressor cells (MDSCs; Fig. 1D). In contrast, the lymphoid cluster (#5) demonstrated downregulation of the T lymphocytes markers (*LY6D, TRBC2*, and *TRBC1*) and the *HLA-DRB1, HLA-DQB1, HLA-C* genes associated with antigenic presentation, while the expression of immunoglobulin markers (*IGHG2, IGHG1, IGLC1*, and *IGKV4-1*) was upregulated (Fig. 1D).

Copy number alteration profile in early-onset TSCC

Functional enrichment analysis of copy number alterations (CNAs) specific to early-onset TSCC revealed PI3K-Akt, MAPK, and metabolic signaling pathways (Supplementary Fig. S1). Three of 5 young patients harbored the same CNAs, including gains on chromosomes 11 (MAPK signaling pathway genes: *ARRB1, MAP3K11, MAP4K2, PAK1, RASGRP2, RELA*, and *RRAS2*) and 20 (TGFb signaling pathway genes: *ID1, TGIF2*, and *RBL1*), and loss on chromosome 5 (*TCF7* gene; Fig. 2).

Fig. 2. Heatmap of copy number alterations in early- and late-onset TSCC. The red-highlighted text indicates recurring copy number alterations where some cancer-related genes are provided. CN, copy number.

Developmental trajectories of tumor cells in early-onset TSCC

Four cases with early-onset TSCC (excluding the FF sample) showed similar differentiation hierarchies with a progenitor cluster of cells (Fig. 3). Progenitor cells were highly differentiated due to the expression of the *KRT16* and *KRTDAP* genes and expressed the MAPK (*IL1B, TNFRSF21, IGF1R, IGFBP3, MAP2K3,* and *MAP2K1*) and tumor growth factor beta (*TGFB1*) pathways genes. The second and third clusters of cases 1 and 4 as well as the second clusters of cases 2 and 3 were decided to characterize together as an intermediate cluster owing to the similarity of gene signature expression. The intermediate cluster originated from progenitor cells was characterized by strong MAPK pathway activation and canonical TGFb signaling inhibition expressing the *DCN, FST,* and *THBS2* genes (inhibitors of the *TGFB1)* and the TGFb signaling activation through the *BMP2* gene. In addition, tumor cells of this cluster demonstrated features of the epithelial-endothelial transition and the activation of vascular mimicry through upregulation of the *EPHA2, FN1, MMP14,* and *COL1A1* genes and the JAK-STAT signaling pathway (*STAT3* and *MYC* genes). The last cluster showed low differentiation due to decreased expression of the *KRT16* and *KRTDAP* genes and retained expression of the MAPK pathway and vascular mimicry genes.

The genes specific to the last cluster of four early-onset TSCC cases were analyzed for association with overall survival using the Human Protein Atlas Kaplan–Meier plotter tool[13]. The focus on the last cluster was related to the suggestion that during the transition of tumor cells from a lower to higher malignant state, some genes are silenced, while others become active, thus predicting aggressiveness and early recurrence of cancer[14,15]. High

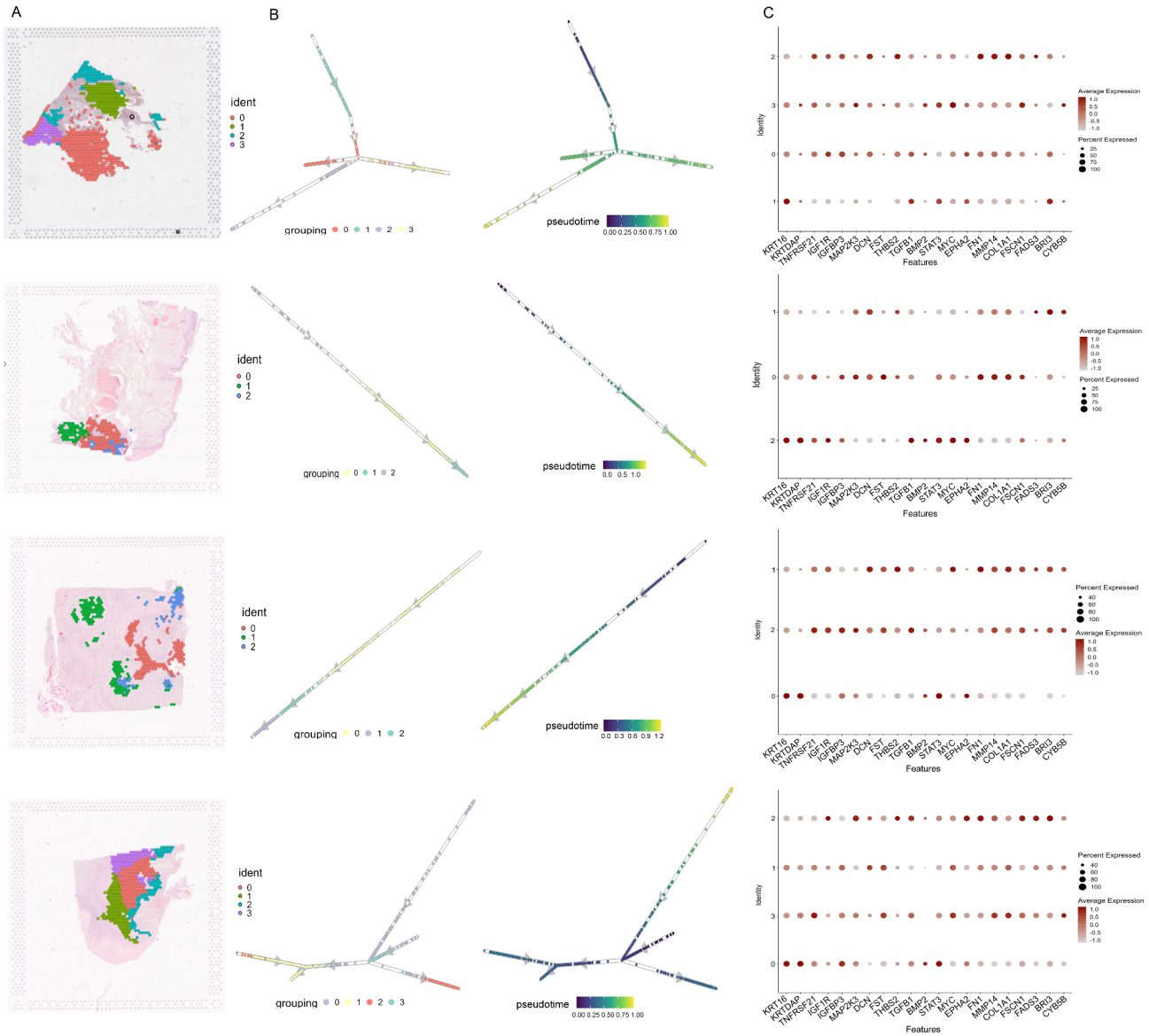

Fig. 3. Tumor cell trajectories in early-onset TSCC and major driver genes. A, Visualization of tumor cell clusters in tissue sections; B, Directions of the cellular trajectories determined by unsupervised pseudo-time; C, Dotplot showing expression of top genes in each tumor cluster timepoint.

expression of the *FSCN1, FADS3, BRI3*, and *CYB5B* genes was found to be associated with worse overall survival of head and neck cancer patients (Supplementary Fig. S1).

Ligand-receptor interactions in the tumor core and invasive front of early-onset TSCC

Ligand-receptor interactions were analyzed in four cases with early-onset TSCC (Fig. 4). The interactions in the fifth young case (FF sample) were described earlier[11]. Among all interactions analyzed, we focused on those whose expression was coincided in most cases. The tumor core demonstrated urokinase-cell surface integrins

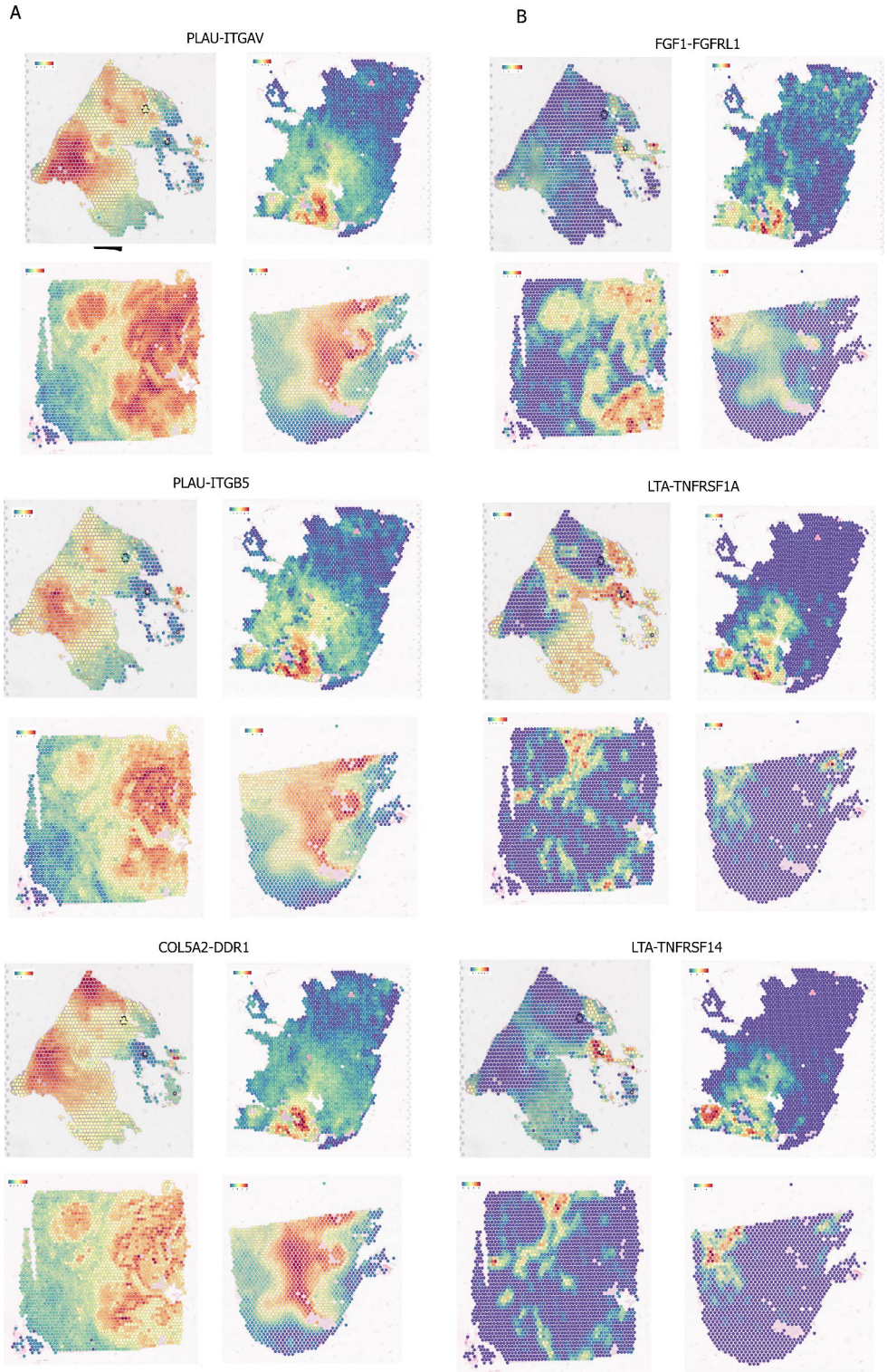

Fig. 4. Ligand-receptor interactions in the tumor core (**A**) and invasive front (**B**) of early-onset TSCC.

(PLAU-ITGAV and PLAU-ITGB5) and collagen-discoidin domain receptor 1 (COL5A2-DDR1) interactions (Fig. 4A). On the invasive front interactions were found between fibroblast growth factor and its receptor (FGF1-FGFRL1) and members of the lymphotoxin alpha and tumor necrosis factor family (LTA-TNFRSF1A and LTA-TNFRSF14; Fig. 4B).

Spatial transcriptome features of the invasive front in early-onset TSCC

In contrast to TSCC in older patients, the border between tumor and normal tissue in 4 of 5 young patients was infiltrated by cells enriched with the immunoglobulin genes such as *IGHG2, IGHG1, IGLC1, IGKV4-1,* and *JCHAIN* corresponding to plasma cells (Fig. 5A). These cells were presented in TLSs, which were absent in the tumors of older patients. The TLS gene signature was also observed in early-onset TSCC (Fig. 5B) and included the above-mentioned immunoglobulin genes and T lymphocytes markers (*TRBC1, TRBC2*, and *IL7R*).

TSCCs of young patients were also distinguished by an increased expression of the *ITGA5, ITGB1, MAP2K3, STAT3, MYC, EPHA2, FN1, MMP14,* and *COL1A1* genes corresponding to the vascular mimicry associated pathway (Fig. 5C). This expression signature was found in solid tumor cell structures located on the border with normal tissue and corresponding to the areas of tumor invasion. Interestingly, regions of vascular mimicry were enriched with macrophage markers: *CD68, SERPINB2*, and *OSM*.

Validation of spatial transcriptomic features of early-onset TSCC by single cell full-length RNA sequencing and TCGA data

To confirm the findings obtained by spatial transcriptomics, we analyzed the TSCC samples from a young patient and an older patient using single cell full-length RNA sequencing (Fig. 6). In addition, the spatial transcriptomics results were validated using the RNA, DNA methylation and protein profiling data of early- and late-onset TSCC presented in the TCGA database (Fig. 6).

Single cell full-length transcriptomes of 26,631 tumor and microenvironment cells have been obtained (Fig. 6A). After basic cell type annotation, we identified 14 immune clusters such as CD4 and CD8 T lymphocytes, and B lymphocytes, T regulatory, natural killer, natural killer T, proliferative T, mast and plasma, and plasmacytoid dendritic cells, as well as 3 TAM clusters (Fig. 6B,C). Immune clusters of early-onset TSCC had decreased expression of T lymphocyte receptor genes such as *TRGC2*, *TRBC1*, and *TRBC2*. Additionally, expression of HLA-DR complex members such as *HLA-DQ1, HLA-DRB5, HLA-DRA1* were reduced in the immune clusters of young patients (Fig. 6E). Also, clusters of tumor, stromal, endothelial, normal epithelium, salivary gland cells, and CAFs were detected (Fig. 6B,C). Similar to spatial transcriptomics findings, the expression of oxidative phosphorylation genes (*COX4I1, NDUFA4, ATP5F1D,* and *ATP5PB*) was decreased, and mitogen activated kinase *MAP3K5* expression was increased in the tumor cells (#10) of young patients compared to older ones (Fig. 6D). In addition to spatial transcriptomics results, we found newly formed endothelial cells (# 8) with high expression of *THY1, COL1A1, LUM,* and lack of *CD34* and *CDH5* genes that were previously described as markers of epithelial-endothelial transition[10,11]. This cluster was distinguished by increased expression of the MAPK pathway regulatory genes: *CACNA1C, RASA2, RASGRP3, RASGRF2*, and *PLA2G4C* (Fig. 6E) and most likely resulted from the transdifferentiation of tumor cells to endothelial cells, the process is known as vascular mimicry[16,17].

Analysis of the TCGA RNA sequencing data showed the downregulation of *KIR2DL4, HLA-DRB5, KLRC1, CD74, HLA-DRB1, HLA-DPA1, HLA-DRA,* and *HLA-DOA* genes associated with antigen processing and presentation in early-onset TSCC compared to older patients (Fig. 6F). Functional enrichment analysis of differentially expressed and differentially methylated genes, as well as differentially expressed proteins, revealed the JAK/STAT and MAPK signaling pathways in tongue tumors of young adults (Fig. 6G). Overall, the single cell full-length RNA sequencing and TCGA data support the transcriptomic features of early-onset TSCC revealed by spatial transcriptomics.

Discussion

The mechanisms of early-onset TSCC are unclear due to a lack of complex investigation of tumor and microenvironment cell compartments. This study used spatial transcriptome sequencing to characterize molecular features of tumor and microenvironment cells in early-onset TSCC. Importantly, we proposed the possibility to integrate FFPE and FF spatial transcriptomics data for combined analysis using established batch correction methods. The spatial transcriptomics findings were confirmed by the single cell full-length RNA sequencing and TCGA data. Unlike older patients, TSCC of young adults was found to be enriched with immunosuppressive gene signature, oxidative stress, the MAPK and JAK-STAT molecular pathways, as well as with plasma cells, TAMs and vascular mimicry features at the invasive front (Fig. 7).

Early-onset TSCC demonstrates extremely downregulated oxidative phosphorylation and upregulated glycolysis, pentose-phosphate pathway, and glutathione metabolism. It is known that imbalance in these systems usually occurs due to oxidative stress and is often observed in cancer[18,19]. This may be due to the rapid growth of the tumor and its aggressiveness. The key molecular player in early-onset TSCC development is most likely the MAPK signaling pathway, which can be activated by immunosuppressive cells and upregulated in response to hypoxia and oxidative stress[18,20,21]. In particular, the MAPK pathway is enriched in tongue tumor cells of young adults compared to older patients. In addition, the tumor core of early-onset TSCC was enriched by the interaction of urokinase (PLAU) with integrins αvβ5, which are known to activate ERK1/2, a key component of the MAPK signaling pathway[22–24]. Activation of the MAPK signaling pathway is known to cause the transformation of normal cells into a tumor phenotype, and its targeting has the potential to arrest tumor growth[25]. Currently, certain MAPK inhibitors are being used in clinical trials, and some of them have led to an improvement in the objective response rate, for example, in non-small cell lung cancer[26]. Non-canonical TGFb signaling pathway through the *BMP2* gene upregulation was also found in early-onset TSCC; however, current data is scarce to

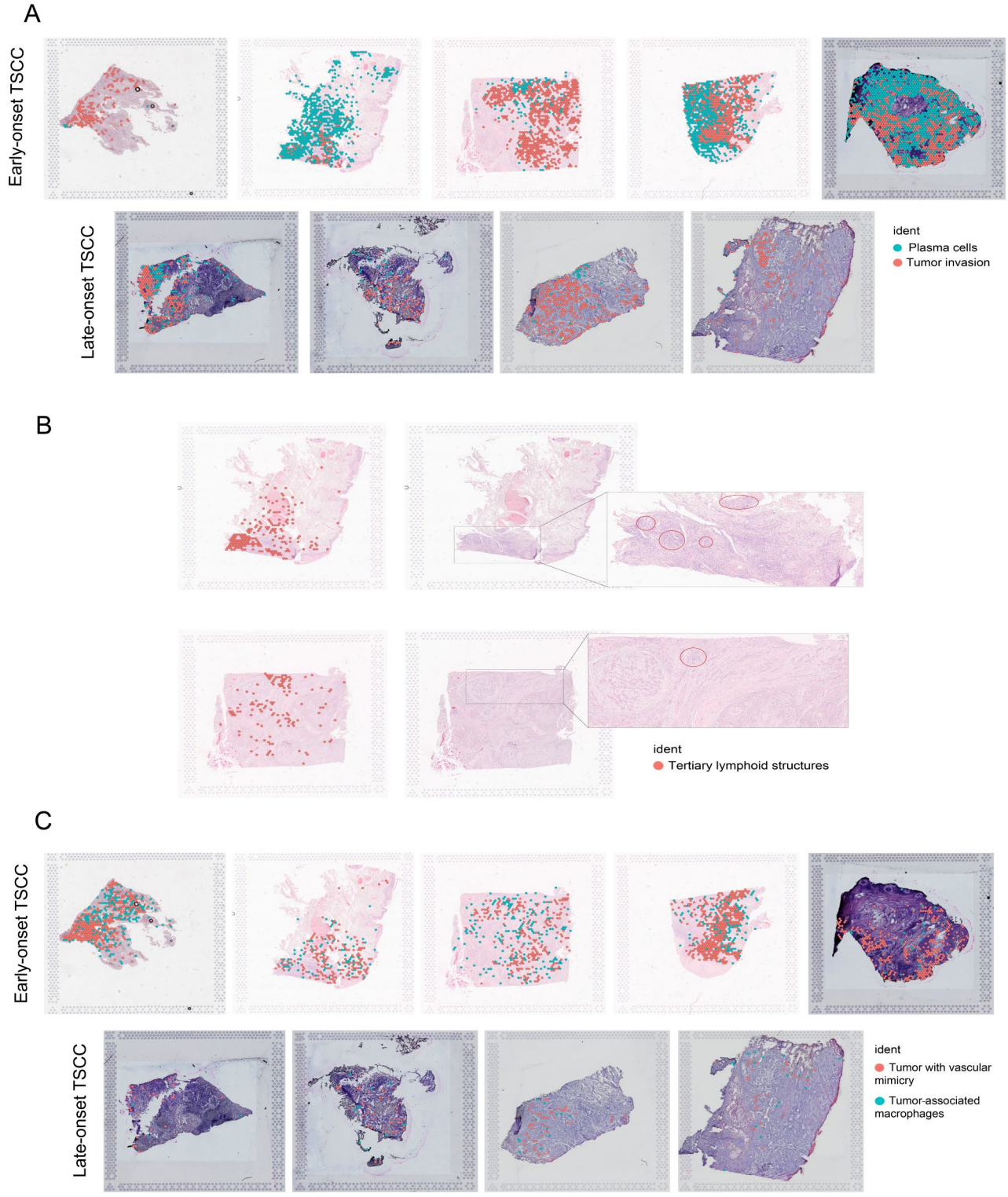

Fig. 5. Spatial transcriptomic features of tumor and microenvironment regions in early- and late-onset TSCC. A, Spatial co-localization of tertiary lymphoid structures with gene signatures of plasma cells (*IGHG2, IGHG1, IGLC1, IGKV4-1*, and *JCHAIN*) and T lymphocytes (*TRBC2, TRBC1*, and *IL7R*); B, Plasma cell gene signature (*IGHG2, IGHG1, IGLC1, IGKV4-1*, and *JCHAIN*) at the border of tumor and normal tissue in early-onset TSCC; C, Markers of vascular mimicry (*ITGA5, ITGB1, MAP2K3, STAT3, MYC, EPHA2, FN1, MMP14*, and *COL1A1*) and TAMs (*CD68, SERPINB2*, and *OSM*) are co-located in solid tumor cell structures at the invasive front.

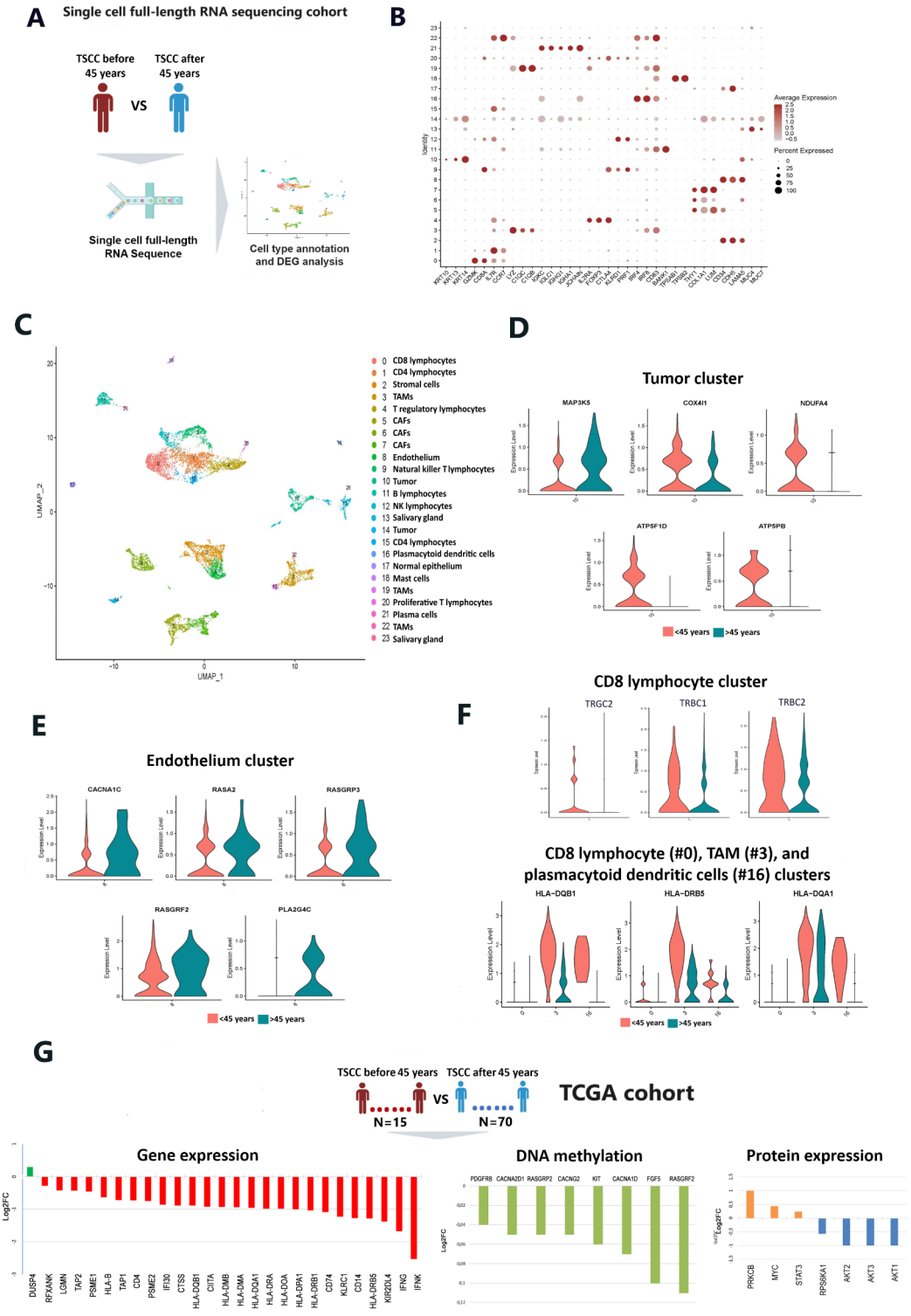

Fig. 6. Transcriptomic features of early-onset TSCC identified via single cell full-length RNA sequencing and TCGA data. A, The experimental design; B, Cluster identification by cell type markers; C, Clustering and cell type annotation of TSCC cells; D-F, Violin plots demonstrating differentially-expressed genes in tumor, endothelial, and immune cell clusters between early- and late-onset TSCC; G, Differentially-methylated and differentially-expressed genes and proteins between early- and late-onset TSCC according to the TCGA database. TSCC, tongue squamous cell carcinoma; DEGs, differentially-expressed genes; CAFs, cancer-associated fibroblasts; TAMs, tumor-associated macrophages; NK, natural killers.

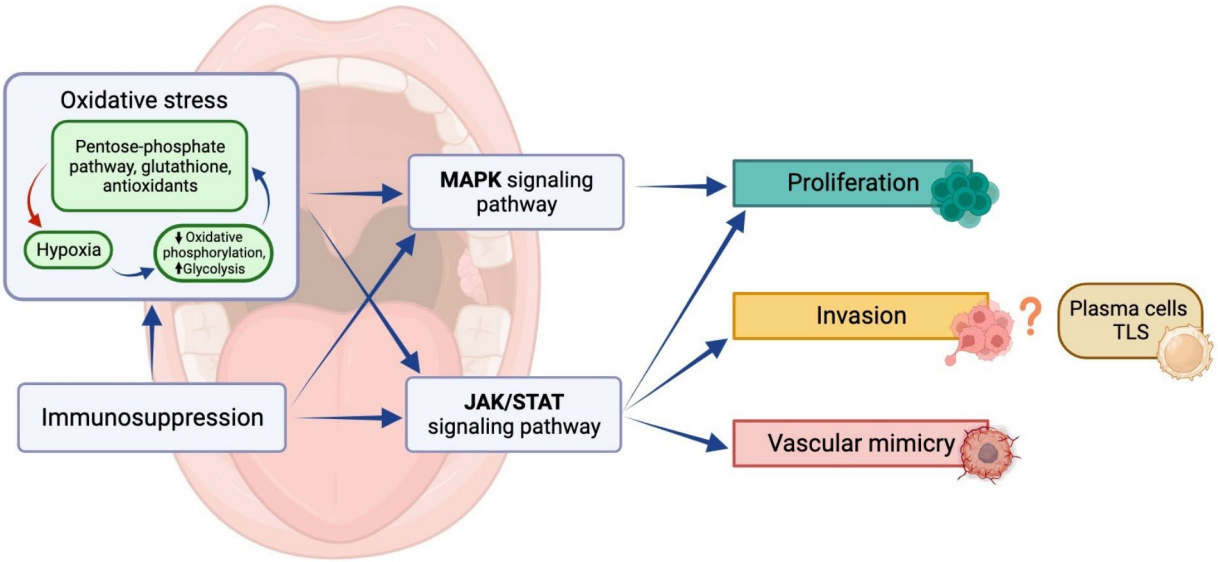

Fig. 7. Hypothetical model of early-onset TSCC pathogenesis. Hypoxia and oxidative stress maintained by immunosuppression seem to be drivers of early-onset TSCC. Oxidative phosphorylation induces tumor development through the MAPK and JAK-STAT signaling pathways that contribute to tumor cell proliferation, apoptosis inhibition, invasion, and vascular mimicry. Plasma cells and TLSs are enriched in the invasive front; however, their role in early-onset TSCC pathogenesis is not yet clear. Red color refers to possible inhibition action; blue color – possible promotion action.

speculate if this signaling is activated[27]. MAPK and TGFb enrichment in early-onset TSCC is most likely related to CNAs on chromosomes 11 and 20 that harbor the genes regulating these signaling pathways. In addition, TSCC in young adults demonstrated the deletion of a region on chromosome 5 affecting the *TCF7* gene, which is the Wnt pathway tumor suppressor gene and a regulator of cell cycle[28].

Tumor microenvironment is well known to contribute to tumorigenesis, cancer progression, and therapy resistance. Many cell types and functionally distinct cell populations comprise the tumor microenvironment demonstrating tumor-supportive or tumor-suppressive roles[29–31]. The early-onset TSCC microenvironment was found to be enriched in MDSCs possessing immunosuppressive and tumor-promoting properties. This cancer also demonstrated a decrease in the expression of the major histocompatibility complex (*HLA-DRB1, HLA-DQB1,* and *HLA-C*) and lymphocyte marker (*LY6L, TRBC2,* and *TRBC1*) genes characterizing an important link in the immune response and antigen presentation. The analysis of the TCGA data also showed PD-L1/PD-1-mediated immunosuppression and inhibition of phagocytosis in TSCC at young age[9]. In light of these findings, we suggest that early-onset TSCC have a pronounced immunosuppression. Moreover, our data calls into question the view that immune checkpoint inhibitors can be effective for treating young patients with TSCC.

The invasive front of the tumor represents an area, where tumor cells actively contact with and invade into surrounding normal tissue contributing to further metastasis and recurrence. The invasive edge of early-onset TSCC is enriched by the LTA-TNFRSF1A, LTA-TNFRSF14, and FGF1-FGFRL1 ligand receptor interactions. LTA is known to be produced by lymphocytes and possess antitumor cytotoxic activity[32]. An interaction between FGF1 and FGFRL1 may indicate invasion through increased phosphorylation of ERK1/2 of the MAPK signaling pathway[33]. In addition, the tumor border is enriched with plasma cells and lymphocytes aggregating to TLSs indicating a better adaptive immune response in young patients as found earlier[34,35]. Nevertheless, as said before, early-onset tongue tumor core is characterized by an increased number of immunosuppressive cells. Further research is needed to reveal the mechanisms of this phenomenon and its importance for disease outcome. The invasive edge of early-onset TSCC was also found to be enriched with vascular mimicry demonstrating upregulation of the JAK-STAT signaling pathway genes and surrounded by TAMs, which may induce or support this process[36–38]. The JAK-STAT signaling pathway was also found to be enriched in TSCC at young age according to the TCGA data[9]. Vascular mimicry or blood vessels formation by tumor cells is often found in solid tumors contributing to tumor cell survival and metastasis[17] and could be a subject of further studies in terms of early-onset cancer.

Conclusion

The present study shows that TSCC in young adults has a distinct spatial transcriptomic features and molecular mechanisms compared to older patients and should be considered a separate clinical form with specific treatment, post-treatment follow-up tactics, and prognosis. Further studies should elucidate the contribution of MAPK and JAK-STAT signaling pathways and local immune defense characteristics to early-onset TSCC in order to develop new diagnostic and prognostic markers and therapeutic targets.

Methods

Tumor sample collection and annotation

The spatial transcriptomic study included primary tumors from two cohorts of patients (n = 9). The clinical data on patients are presented in Supplementary Table 1. The first cohort included 4 patients (aged < 45 years) with TSCC treated at the Cancer Research Institute of Tomsk National Research Medical Center of the Russian Academy of Sciences. The second cohort included 5 patients (1 patient aged < 45 years and 4 patients aged > 45 years), tumor samples of which were obtained from the Alberta Cancer Research Biobank. All patients did not receive neoadjuvant treatment. Tumor samples from the first cohort were FFPE, and the second ones were FF. The group of 5 patients aged up to 45 years (4 FFPE and 1 FF samples) was compared with the group of 4 patients aged over 45 years (4 FF samples) (Fig. 1A).

Single cell full-length RNA sequencing was performed using FF samples of TSCC obtained during surgery of 1 patient aged < 45 years and 1 patient aged > 45 years. The clinical data on these patients are presented in Supplementary Table 1. All patients did not receive neoadjuvant treatment. Tumor samples were stored in RPMI 1640 (Gibco, USA) containing 10% FBS on ice before preparation of single-cell suspension.

Spatial transcriptomic profiling

Visium spatial gene expression slides and reagents (Visium Human Transcriptome Probe Set v1.0, 10 × Genomics, USA) for spatial transcriptomics assay were used according to manufacturer instructions. To assess the quality of the selected FFPE blocks, RNA was isolated by PureLink FFPE RNA Isolation Kit (Invitrogen, USA) from serial 10 µm-thick tissue sections. Determination value index DV_{200} that represents the percentage of RNA fragments larger than 200 nucleotides with respect to all RNA fragments was calculated using the 4150 TapeStation system (Agilent, USA). Samples which had a $DV_{200} \geq 50\%$ were considered good quality and selected to proceed with the experiment. Five-micrometer FFPE sections of TSCC samples were placed on a Visium Gene Expression slide (10 × Genomics, USA). Deparaffinization, hematoxylin and eosin staining, and decrosslinking were performed according to the protocol of Visium Spatial Gene Expression for FFPE (10 × Genomics, USA). Leica Aperio AT2 histoscanning station (Leica, Germany) was used to visualize tissue sections on slides. After incubation with the Probe Hybridization Mix (10 × Genomics, USA), the tissues were permeabilized and captured. Gene Expression (GEX) libraries were generated for each section and then sequenced on a GenoLab M (GeneMind Biosciences, China). The Visium data on FF samples were kindly provided by Pinaki Bose and Rohit Arora[11].

Spatial transcriptomics data processing

Sequencing reads were aligned using the 10 × Genomics Space Ranger 1.3.1 for FF tissues and 2.0.0. pipelines (10 × Genomics, USA) for FFPE tissues to the standard GRCh38 reference genome. Seurat package (v.4.3.0) was used for the following analysis. For each sample, the spatial file and H5 file obtained from SpaceRanger were loaded by the function "Load10X_spatial". We filtered each Seurat Objects to exclude low-quality and outlier spots. The filtering parameters are presented in Supplementary Table 2. For spatial transcriptomics data obtained from FF samples, mitochondrial genes were excluded from the analysis. Data were normalized using the "NormalizeData" function with a scale factor of 10,000. Variable genes were identified using the "FindVariableFeatures" function. Data were scaled and centered using the "ScaleData" function. Harmony package version 0.1.1 was used to integrate data correcting by material type (FF or FFPE). For optimal results, genes not captured by FFPE probes were removed from the FF spatial transcriptomics data prior to integration. Cell clusters were identified via the "FindNeighbors" and "FindClusters" functions using a resolution of 0.5 for integrated object. "FindAllMarkers" function was used to find differentially expressed genes between all clusters and according to patient group. We used a Wilcoxon rank-sum test with a logFC > 0.5 and an adjusted p value < 0.001.

Spatial transcriptomic cell type annotation

Marker genes were used to determine whether clusters belonged to known cell types: CAFs – *COL1A1, VIM,* and *LUM*; tumor cells – *KRT1, KRT4, KRT5, KRT10, KRT13,* and *KRT1;* myocytes – *TTN* and *DES;* salivary glands cells – *MUC;* lymphoid cells – *IGHG2, IGHG1, IGLC1, IGHM, IGHJ6, CCL19, CD3D, CD3E,* and *CD4;* myeloid cells – *SERPINEB2, LYZ, TREM1, FCGR2A,* and *OSM* (Fig. 1C). The "DotPlot" function (Seurat package) was used to visualize the expression of marker genes across different clusters (Fig. 1C).

CNA inference

CNAs were analyzed using inferCNV (https://github.com/broadinstitute/infercnv) for each sample independently using microenvironment spots as a reference. For sample 4 from the young patient group, normal epithelium spots were used as a reference. The inferCNV analysis included the following parameters: cutoff = 0.1, denoise = TRUE, noise_filter = 0.1, HMM = TRUE. Pathway and functional enrichment analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG) was performed using the Enrichr[39–41].

Trajectory inference

After isolating tumor cell spots with minimal stromal cell content from young patient samples, each sample was independently used for maturation trajectory inference using dynverse packages, which provide more than 50 different methods for trajectory inference. Normalized tumor cell data and UMAP coordinates were used as input. We also entered prior information about start cell identifier, which was chosen based on the expression of the genes responsible for epithelial cell differentiation (*KRTDAP*, *MKI67*, *CD44*, and *S100A14*), as well as information about the *KRTDAP* gene that is known to be important in oral cell differentiation[42]. The most suitable and rapid method for our dataset determined by applying "guidelines" function was the PAGA Tree. The visualization of the trajectories and pseudotime as a 2D graph structure was done with the "plot_graph" function

(dynplot package). A standard "SpatialDimPlot" function (Seurat package) was used to visualize clusters of tumor cells on slides.

Tumor core and invasive front determination

Morphologically, the tumor core was identified as multi-sized nests and large solid tumor fields with dense intercellular contacts and tissue stroma without signs of edema and inflammation. The invasive front was determined if small nests and clusters of tumor cells were located in loose fibrous connective tissue stroma with signs of desmoplasia, edema, and moderate inflammatory infiltration.

Ligand-receptor interaction analysis

NICHES R package was used to assess ligand-receptor interactions between neighbor tissue spots in spatial transcriptomics data[43]. Adaptively-thresholded Low Rank Approximation (ALRA) imputation on log normalized counts was used to mitigate drop-out effect[44]. ALRA-imputed gene expressions were then used in NICHES analysis. "FindAllMarkers" function in Seurat was used to reveal most significant ligand-receptor interactions in the individual ligan-receptor interaction clusters. Visualization of results was performed using Seurat package. Receiver operating characteristic (ROC) analysis was used to compare ligand-receptor interactions in the tumor core (Area under the curve (AUC) > 0.85) and invasive front (AUC > 0.75).

Single cell full-length RNA sequencing

Fat tissues and visible blood vessels were removed before tissue processing. TSCC tissues were washed with ice-cold PBS and dissociated using the Tumor Dissociation Kit (human) and a gentleMACS Octo Dissociator with heaters (Miltenyi, Germany). Subsequent steps were performed at 4 °C. Cell suspensions were filtered using 70 µm filters, pelleted by centrifugation at 300 × g in BSA-coated low-binding tubes, treated with 1 ml VersaLyse Lysing Solution (Beckman Coulter, USA) for 5 min, washed with RPMI (Gibco, USA), pelleted, resuspended in PBS, and filtered using 40 µm filters. Cell viability was estimated using flow cytometry with the CytoFLEX instrument (Beckman Coulter, USA), which displayed a viability rate exceeding 90% of Draq7 (BD Bioscience, USA) negative cells.

The cell suspension was further diluted to a concentration of 1000 cells/µl. Single cell full-length RNA sequencing libraries were prepared from two tumor samples of young and older patients using the SeekOne DD Single Cell Full-length RNA Sequence Transcriptome-seq Kit (Beijing SeekGene BioSciences, China). Libraries were quantified and qualified by the Qubit 4 fluorometer (Thermo Fisher Scientific, USA) and 4150 TapeStation system (Agilent, USA), and sequenced on the Genolab M (GeneMind Biosciences, China).

Single cell full-length RNA sequencing data bioinformatics

Sequencing data were aligned to the human reference genome (GRCh38) and processed using seeksoultools.1.2.0 (Beijing SeekGene BioSciences, China). Seurat package (v. 4.3.0) was used for the following analysis. We filtered each Seurat Objects to exclude low-quality and outlier spots: cells with detected gene counts between 2000 and 7500 (for an older patient) and between 500 and 6000 (for a young patient) and mitochondrial gene fraction less than 25%, and mitochondrial genes were removed from the expression table. Data were normalized using the "SCTransform" after PCA dimensionality reduction. Harmony package version 0.1.1 was used for data integration. Cell clusters were identified via the "FindNeighbors" and "FindClusters" functions, using a resolution of 0.8 per object. "FindAllMarkers" function was used to find differentially expressed genes between all clusters and according to patient group. We used a Wilcoxon rank-sum test with a logFC > 0.25 and an adjusted p value < 0.001.

TCGA data analysis

A human tongue cancer dataset was downloaded from The Cancer Genome Atlas database (https://www.cancer.gov/tcga) through cBioPortal (https://www.cbioportal.org). The TCGA cohort contained 85 tumor samples with survival and follow-up information of TSCC patients, who were divided into two groups: 15 patients (aged < 45 years) and 70 patients (aged > 45 years). The clinical data of patients are presented in Supplementary Table 1. All data were obtained from publicly available databases and did not require approval of the local ethics committee. Identification of differentially expressed genes was performed using the DESeq2 package in R with log2FC ≥ 0.75 and log2FC ≤ −0.75 and adjusted p value < 0.05. DNA methylation profile and protein expression were analyzed using cBioPortal online tool. Student's t-test with log2FC > 0 and log2FC < 0 and p value < 0.05 was used for the analysis of protein expression and DNA methylation profile of TSCC according to the patient's age. KEGG pathway and functional enrichment analysis was performed using the Enrichr[39–41].

Statistical analysis

The Human Protein Atlas Kaplan–Meier plotter tool was used to evaluate the correlation between the TCGA gene expression and the survival rates in 499 head and neck cancer patients[45]. The total dataset included 133 females and 366 males. Around 56% of the patients (281 patients) were still alive at the time of data collection. The stage distribution was stage I 25 patients, stage II 69 patients, stage III 78 patients, stage IV 259 patients and 68 patients with missing stage information.

Data availability

Data is provided within the manuscript or supplementary information files.

Received: 22 March 2024; Accepted: 10 October 2024

Published online: 01 November 2024

References

1. da Silva Souto, A. C. et al. Epidemiology of tongue squamous cell carcinoma: A retrospective cohort study. *Oral Dis.* **29**, 402–410. https://doi.org/10.1111/odi.13897 (2023).
2. Chen, J. K., Eisenberg, E., Krutchkoff, D. J. & Katz, R. V. Changing trends in oral cancer in the United States, 1935 to 1985: A Connecticut study. *J. Oral Maxillofac. Surg. Off. J. Am. Assoc. Oral Maxillofac. Surg.* **49**, 1152–1158. https://doi.org/10.1016/0278-2391(91)90406-c (1991).
3. Kolegova, E. S. et al. Early-onset oral cancer as a clinical entity: Aetiology and pathogenesis. *Int. J. Oral Maxillofac. Surg.* **51**, 1497–1509. https://doi.org/10.1016/j.ijom.2022.04.005 (2022).
4. Nishizawa, R. et al. The 2G allele of promoter region of matrix metalloproteinase-1 as an essential pre-condition for the early onset of oral squamous cell carcinoma. *BMC Cancer* **7**, 187. https://doi.org/10.1186/1471-2407-7-187 (2007).
5. Dutta, A., Saikia, N., Phookan, J., Baruah, M. N. & Baruah, S. Association of killer cell immunoglobulin-like receptor gene 2DL1 and its HLA-C2 ligand with family history of cancer in oral squamous cell carcinoma. *Immunogenetics* **66**, 439–448. https://doi.org/10.1007/s00251-014-0778-1 (2014).
6. Campbell, B. R. et al. The mutational landscape of early- and typical-onset oral tongue squamous cell carcinoma. *Cancer* **127**, 544–553. https://doi.org/10.1002/cncr.33309 (2021).
7. D'Souza, W. & Saranath, D. OMICS, oral cancer molecular landscapes, and clinical practice. *Omics J. Integr. Biol.* **21**, 689–703. https://doi.org/10.1089/omi.2017.0146 (2017).
8. Maroun, C. A. et al. An immunogenomic investigation of oral cavity squamous cell carcinoma in patients aged 45 years and younger. *The Laryngoscope* **131**, 304–311. https://doi.org/10.1002/lary.28674 (2021).
9. Prostakishina, E. A. et al. molecular landscape of oral cancer in young adults. *Russ. J. Genet.* **59**, 1190–1201. https://doi.org/10.1134/S1022795423110108 (2023).
10. Sun, L. et al. Single-cell and spatial dissection of precancerous lesions underlying the initiation process of oral squamous cell carcinoma. *Cell Discov.* **9**, 28. https://doi.org/10.1038/s41421-023-00532-4 (2023).
11. Arora, R. et al. Spatial transcriptomics reveals distinct and conserved tumor core and edge architectures that predict survival and targeted therapy response. *Nat. Commun.* **14**, 5029. https://doi.org/10.1038/s41467-023-40271-4 (2023).
12. Tran, H. T. N. et al. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* **21**, 12. https://doi.org/10.1186/s13059-019-1850-9 (2020).
13. Karlsson, M. et al. A single-cell type transcriptomics map of human tissues. *Sci. Adv.* https://doi.org/10.1126/sciadv.abh2169 (2021).
14. Qin, S. et al. Emerging role of tumor cell plasticity in modifying therapeutic response. *Signal transduct. Target. Ther.* **5**, 228. https://doi.org/10.1038/s41392-020-00313-5 (2020).
15. Rozenberg, J. M. et al. Molecules promoting circulating clusters of cancer cells suggest novel therapeutic targets for treatment of metastatic cancers. *Front. Immunol.* **14**, 1099921. https://doi.org/10.3389/fimmu.2023.1099921 (2023).
16. Maddison, K., Bowden, N. A., Graves, M. C. & Tooney, P. A. Characteristics of vasculogenic mimicry and tumour to endothelial transdifferentiation in human glioblastoma: A systematic review. *BMC Cancer* **23**, 185. https://doi.org/10.1186/s12885-023-10659-y (2023).
17. Wechman, S. L., Emdad, L., Sarkar, D., Das, S. K. & Fisher, P. B. Vascular mimicry: Triggers, molecular interactions and in vivo models. *Adv. Cancer Res.* **148**, 27–67. https://doi.org/10.1016/bs.acr.2020.06.001 (2020).
18. Di Gregorio, J., Petricca, S., Iorio, R., Toniato, E. & Flati, V. Mitochondrial and metabolic alterations in cancer cells. *Eur. J. Cell Biol.* **101**, 151225. https://doi.org/10.1016/j.ejcb.2022.151225 (2022).
19. Abdel-Wahab, A. F., Mahmoud, W. & Al-Harizy, R. M. Targeting glucose metabolism to suppress cancer progression: Prospective of anti-glycolytic cancer therapy. *Pharmacol. Res.* **150**, 104511. https://doi.org/10.1016/j.phrs.2019.104511 (2019).
20. Brägelmann, J. et al. MAPK-pathway inhibition mediates inflammatory reprogramming and sensitizes tumors to targeted activation of innate immunity sensor RIG-I. *Nat. Commun.* **12**, 5505. https://doi.org/10.1038/s41467-021-25728-8 (2021).
21. Cheng, Y., Chen, J., Shi, Y., Fang, X. & Tang, Z. MAPK signaling pathway in oral squamous cell carcinoma: Biological function and targeted therapy. *Cancers* **14**, 4625. https://doi.org/10.3390/cancers14194625 (2022).
22. Mekkawy, A. H., Pourgholami, M. H. & Morris, D. L. Involvement of urokinase-type plasminogen activator system in cancer: An overview. *Med. Res. Rev.* **34**, 918–956. https://doi.org/10.1002/med.21308 (2014).
23. Alfano, D., Franco, P. & Stoppelli, M. P. Modulation of cellular function by the Urokinase receptor signalling: A mechanistic view. *Front. Cell Dev. Biol.***10**, 818616. https://doi.org/10.3389/fcell.2022.818616 (2022).
24. Franco, P. et al. Activation of urokinase receptor by a novel interaction between the connecting peptide region of urokinase and alpha v beta 5 integrin. *J. Cell Sci.* **119**, 3424–3434. https://doi.org/10.1242/jcs.03067 (2006).
25. Sun, Y. et al. Signaling pathway of MAPK/ERK in cell proliferation, differentiation, migration, senescence and apoptosis. *J. Recept. Signal Transduct. Res.* **35**, 600–604. https://doi.org/10.3109/10799893.2015.1030412 (2015).
26. Carter, C. A. et al. Selumetinib with and without erlotinib in KRAS mutant and KRAS wild-type advanced nonsmall-cell lung cancer. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **27**, 693–699. https://doi.org/10.1093/annonc/mdw008 (2016).
27. Yang, Y. et al. The role of TGF-β signaling pathways in cancer and its potential as a therapeutic target. *Evidence-Based Complement. Altern. Med. eCAM* **2021**, 6675208. https://doi.org/10.1155/2021/6675208 (2021).
28. Jung, K. H. et al. Loss-of-function mutations in the Transcription Factor 7 (T cell factor-1) gene in hepatogastrointestinal cancers. *Mol. Cell. Toxicol.* **6**, 271–278. https://doi.org/10.1007/s13273-010-0037-y (2010).
29. Hanahan, D. & Coussens, L. M. Accessories to the crime: Functions of cells recruited to the tumor microenvironment. *Cancer Cell* **21**, 309–322. https://doi.org/10.1016/j.ccr.2012.02.022 (2012).
30. Anderson, N. M. & Simon, M. C. The tumor microenvironment. *Current Biol. CB* **30**, R921-r925. https://doi.org/10.1016/j.cub.2020.06.081 (2020).
31. Dzobo, K., Senthebane, D. A. & Dandara, C. The tumor microenvironment in tumorigenesis and therapy resistance revisited. *Cancers* **15**, 376. https://doi.org/10.3390/cancers15020376 (2023).
32. Gubernatorova, E. O. et al. Dual role of TNF and LTα in carcinogenesis as implicated by studies in mice. *Cancers* **13**, 1775. https://doi.org/10.3390/cancers13081775 (2021).
33. Francavilla, C. et al. The binding of NCAM to FGFR1 induces a specific cellular response mediated by receptor trafficking. *J. Cell Biol.* **187**, 1101–1116. https://doi.org/10.1083/jcb.200903030 (2009).
34. Phanthunane, C. B-cell clusters at the invasive margin associate with longer survival in early-stage oral-tongue cancer patients. *Oncoimmunology* **10**, 1882743. https://doi.org/10.1080/2162402x.2021.1882743 (2021).
35. Meylan, M. Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer. *Immunity* **55**, 527-541.e525. https://doi.org/10.1016/j.immuni.2022.02.001 (2022).
36. Rong, X. et al. Tumor-associated macrophages induce vasculogenic mimicry of glioblastoma multiforme through cyclooxygenase-2 activation. *Oncotarget* **7**, 83976–83986. https://doi.org/10.18632/oncotarget.6930 (2016).
37. Liu, Q. et al. Tumor-associated macrophage-derived exosomes transmitting miR-193a-5p promote the progression of renal cell carcinoma via TIMP2-dependent vasculogenic mimicry. *Cell Death Dis.* **13**, 382. https://doi.org/10.1038/s41419-022-04814-9 (2022).

38. Qiao, T. et al. Dynamic differentiation of F4/80+ tumor-associated macrophage and its role in tumor vascularization in a syngeneic mouse model of colorectal liver metastasis. *Cell Death Dis.* **14**, 117. https://doi.org/10.1038/s41419-023-05626-1 (2023).
39. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951. https://doi.org/10.1002/pro.3715 (2019).
40. Kanehisa, M. et al. KEGG for taxonomy-based analysis of pathways and genomes. *Nucl. Acids Res.* **51**, 587-D592. https://doi.org/10.1093/nar/gkac963 (2023).
41. Kuleshov, M. V. et al. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucl. Acids Res.* **44**, W90-97. https://doi.org/10.1093/nar/gkw377 (2016).
42. Su, Y. et al. PSMC2, ORC5 and KRTDAP are specific biomarkers for HPV-negative head and neck squamous cell carcinoma. *Oncol. Lett.* **21**, 289. https://doi.org/10.3892/ol.2021.12550 (2021).
43. Raredon, M. S. B. et al. Comprehensive visualization of cell-cell interactions in single-cell and spatial transcriptomics with NICHES. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btac775 (2023).
44. Linderman, G. C. et al. Zero-preserving imputation of single-cell RNA-seq data. *Nat. Commun.* **13**, 192. https://doi.org/10.1038/s41467-021-27729-z (2022).
45. Uhlen, M. et al. A pathology atlas of the human cancer transcriptome. *Science* **357**, 6352. https://doi.org/10.1126/science.aan2507 (2017).

Author contributions

MRP, ESK, ELC, and EVD contributed to study conception and design. MRP, EAP, MEM, IVL, OIK, DML, MVZ, IKF, DEK, and RA contributed to developing methodology. APP, LPY, MAK, and NSS contributed to data acquisition (managed patients, provided facilities, etc.). AAK, VYK, PSI, and RA contributed to data analysis and interpretation (e.g., statistical analysis, biostatistics, and computational analysis). YL, LJ, ELC, PB and EVD contributed to writing, review, and/or revision of the manuscript. ESK, EAP, IKF and DEK contributed to administrative, technical, or material support (e.g., reporting or organizing data and constructing databases). ELC and EVD contributed to study supervision. All authors read and approved the final manuscript. The Consortium Etiology and Pathogenesis of Oral Cancer in Young Adults members collected the data.

Funding

This work was supported by the Russian Science Foundation (# 22-15-00308).

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

The study was approved by the Ethics Committee of the Cancer Research Institute of Tomsk National Research Medical Center of the Russian Academy of Sciences (Study ID: No. 12 of 15 June 2022) and Health Research Ethics Board of Alberta – Cancer Committee (Study ID: HREBA.CC-16–0644) and performed according to the guidelines of Declaration of Helsinki and the International Conference on Harmonization's Good Clinical Practice Guidelines (ICH GCP). All subjects gave written informed consent in accordance with the Declaration of Helsinki.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-76044-2.

Correspondence and requests for materials should be addressed to E.S.K.

The Consortium Etiology and Pathogenesis of Oral Cancer in Young Adults

Marina R. Patysheva[1], Elena S. Kolegova[1], Elizaveta A. Prostakishina[1], Vyacheslav Y. Korobeynikov[1], Irina K. Fedorova[5], Denis E. Kulbakin[5], Andrey P. Polyakov[7], Liliya P.

Yakovleva[8], Mikhail A. Kropotov[9], Natalya S. Sukortseva[10], Evgeny L. Choinzonov[5] & Evgeny V. Denisov[1,2]