



## OPEN ACCESS

## EDITED BY

Xin Jiang,  
First Affiliated Hospital of Jilin University, China

## REVIEWED BY

Zhe Wang,  
Affiliated Zhongshan Hospital of Dalian  
University, China  
Yingming Sun,  
Fujian Medical University, China  
Martin Kauke-Navarro,  
Yale-New Haven Hospital, United States

## \*CORRESPONDENCE

Qiao Qiao,  
✉ braveheart8063@outlook.com  
Xuexin Tan,  
✉ tan.xue.xin@hotmail.com

†These authors have contributed equally to  
this work

RECEIVED 24 November 2023

ACCEPTED 18 June 2024

PUBLISHED 12 July 2024

## CITATION

Qi K, Li G, Jiang Y, Tan X and Qiao Q (2024),  
Stromal cell-expressed malignant gene  
patterns contribute to the progression of  
squamous cell carcinomas across  
different sites.  
*Front. Genet.* 15:1342306.  
doi: 10.3389/fgene.2024.1342306

## COPYRIGHT

© 2024 Qi, Li, Jiang, Tan and Qiao. This is an  
open-access article distributed under the terms  
of the [Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in  
other forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in this  
journal is cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Stromal cell-expressed malignant gene patterns contribute to the progression of squamous cell carcinomas across different sites

Kaiyan Qi<sup>1†</sup>, Guangqi Li<sup>2†</sup>, Yuanjun Jiang<sup>3</sup>, Xuexin Tan<sup>4\*†</sup> and  
Qiao Qiao<sup>1\*†</sup>

<sup>1</sup>Department of Radiation Oncology, The First Hospital of China Medical University, Shenyang, China, <sup>2</sup>Department of Biotherapy, Cancer Center and State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, China, <sup>3</sup>Department of Urology, The First Hospital of China Medical University, Shenyang, China, <sup>4</sup>Department of Oral Maxillofacial-Head and Neck Surgery, School and Hospital of Stomatology, China Medical University, Shenyang, China

**Background:** Squamous cell carcinomas (SCCs) across different anatomical locations possess common molecular features. Recent studies showed that stromal cells may contribute to tumor progression and metastasis of SCCs. Limited by current sequencing technology and analysis methods, it has been difficult to combine stroma expression profiles with a large number of clinical information.

**Methods:** With the help of transfer learning on the cell line, single-cell, and bulk tumor sequencing data, we identified and validated 2 malignant gene patterns (V1 and V5) expressed by stromal cells of SCCs from head and neck (HNSCC), lung (LUSC), cervix (CESC), esophagus, and breast.

**Results:** Pattern V5 reflected a novel malignant feature that explained the mixed signals of HNSCC molecular subtypes. Higher expression of pattern V5 was related to shorter PFI with gender and cancer-type specificity. The other stromal gene pattern V1 was associated with poor PFI in patients after surgery in all the three squamous cancer types (HNSCC  $p = 0.0055$ , LUSC  $p = 0.0292$ , CESC  $p = 0.0451$ ). Cancer-associated fibroblasts could induce HNSCC cancer cells to express pattern V1. Adjuvant radiotherapy may weaken the effect of high V1 on recurrence and metastasis, depending on the tumor radiosensitivity.

**Conclusion:** Considering the prognostic value of stromal gene patterns and its universality, we suggest that the genetic subtype classification of SCCs may be improved to a new system that integrates both malignant and non-malignant components.

## KEYWORDS

squamous cell carcinoma, head and neck, stromal cell, cancer associated fibroblast, lung cancer

## Background

Squamous cell carcinomas (SCCs) are among the most prevalent forms of solid cancers, originating from epithelia of multiple organs, such as head and neck, esophagus, lung, skin, and cervical. SCCs across different anatomical locations possess common histopathological and molecular features (Schwaederle et al., 2015; Chai et al., 2020), showing a high tendency

to relapse and metastasize. Although there has been substantial progress in surgery, radiotherapy, and targeted therapy of these tumors, the prognosis remains stagnant due to locoregional recurrences, second primary tumors, and distant metastases.

There is increasing evidence that tumors result from disordered organ homeostasis rather than malignant single cell (or cells). A landmark study on SCCs of head and neck (HNSCC) first put forward the theory of field cancerization (Slaughter et al., 1953) to explain the surprisingly high frequency of multifocality and second primary tumors in epithelial cancers (Rahal et al., 2022). In the first step of field cancerization, genetically altered epithelial proliferate to form a field lesion without any histological changes. Then certain cells within this field got further critical changes and transformed into malignant cells, resulting in multi-foci primary tumors (changes occur at the same time) or second primary tumors (changes occur in sequence) (Baumeister et al., 2021). The theory has been accepted in numerous epithelial cancers, drawing attention to normal-appearing adjacent tissue surrounding the tumor. However, in this model, changes of the epithelium are primary determinants, while the stromal alterations received relatively less attention. As researches on cancer drivers shifting from cancer cells to tumor microenvironment (TME), alteration in tumor stroma has been found to play a primary role in carcinogenesis (Bhowmick et al., 2004; Dotto, 2014). Many cancer driver mutations can be found in tumor stroma (Guan et al., 2020), which shows different structures from normal stroma (Plava et al., 2019; Xu et al., 2022). Hypotheses have been proposed that stroma could contribute to field cancerization (Cox, 2021), and researches have shown the contribution of stromal alteration to cancer progression and metastasis (Hanahan and Weinberg, 2011; Valkenburg et al., 2018; Chhabra and Weeraratna, 2023).

Although the deconvolution technique and single-cell sequencing are under rapid development, it is still hard to illustrate cancer stromal alterations across different types of transcriptomic data. Previous researches on the relationship between the genetic features and tumor prognosis of SCCs were limited to single anatomical locations, focusing on cancer cells (or mixed signals) rather than stromal cells. One reason is that bulk sequencing loses the signal complexity within a tumor while single-cell sequencing compromises on the number of tumors sequenced, therefore it has been hard to combine stroma expression profiles with a large number of prognostic information. Therefore, previous sequencing-based studies on cancer stroma focused on stromal cell proportion rather than gene alteration of stromal cells. To solve this problem, we designed a series of analytical approaches based on matrix factorization, transfer learning and deconvolution for cell line, single-cell, and bulk tumor sequencing data. We recognized squamous malignant gene patterns across different organs and projected the patterns to specific datasets to explore the relationship between gene patterns and phenotypes. Focusing on the cancer stroma, we explored the prognostic value of the stromal-expressed malignant gene patterns and their contribution to subtype classification of SCCs.

## Methods

### Datasets

The microarray gene expression profiles of the 63 squamous cancer cell lines were selected from the Sanger Cell Line Affymetrix

Gene Expression Project (GSE68950) (NCHBI, 2015). The single-cell RNA sequencing (scRNA-seq) data of head and neck squamous cancer (HNSCC) has been published previously (Puram et al., 2017). Single-cell transcriptomes for 5,902 cells from 18 HNSCC patients with both more than 2,000 detected genes and an average housekeeping expression level above 2.5 passed initial quality controls. RNA-seq, DNA methylation, and clinical data of TCGA HNSCC, lung squamous cell carcinoma (LUSC), and cervical squamous cell carcinoma (CESC) datasets were obtained from Genomic Data Commons Data Portal (National Cancer Institute, 2024). The 303 reference methylation profiles of the known cell types used for deconvolution were gathered from the Gene Expression Omnibus (GEO) (NCHBI, 2023). The microarray gene expression profiles of five HNSCC cell lines grown without and with patient-matched cancer-associated fibroblasts (CAFs) were obtained from GSE178153 (NCHBI, 2021). The 98 paired normal and HNSCC tumor samples used in Figure 3H were selected from (Smetana et al., 2020).

### Pattern identification and transfer learning

The nonnegative matrix factorization (NMF) has been applied for multiple purposes including image processing, language modeling, and genomic feature extraction (Devarajan, 2008). This approach uses a limited number of basic components to interpret the target data as accurately as possible. We performed NMF algorithm (Gaujoux and Seoighe, 2010) in a relatively heterogeneous dataset of multiple squamous cancer types to recognize patterns associated with squamous malignancy. Then we projected the patterns to specific datasets to explore the relationship between gene patterns and phenotypes. For example, we transferred the learned patterns to a single cell transcriptome to explore their distribution in different cell types. This transfer learning process was performed using the R package “projectR” (23). Pattern expression levels was defined as projection score.

### Deconvolution and cell-type gene expression estimation

Deconvolution technology for bulk tissue sequencing aims to fill the gap where bulk sequencing loses the signal complexity within a tumor while single-cell sequencing compromises on the number of tumors sequenced. In many cases where a sufficient number of patients is needed to satisfy the statistical test, deconvolution is the best way if we want to keep the sequencing accuracy at the level of cell types rather than bulk tissue.

The Edec R package (Onuchic et al., 2016) was used for deconvolution. A total of 303 GEO DNA methylation profiles of known cell types (cancer cells, stromal cells, and immune cells, Supplementary Table S1) were collected as reference (Li et al., 2021). First, we identified 400 DNA loci from the 450 k methylation profile which allowed us to accurately distinguish the three reference cell types (Supplementary Figure S1A). Based on the 400 loci, we deconvoluted the TCGA HNSCC methylation profile into three subtypes. The three parts showed a high correlation with cancer, stroma, and immune reference profiles, respectively (Supplementary

Figure S1B). We validated the deconvolution result by comparing the cancer cell type proportion with the cancer purity calculated by the ABSOLUTE method (Carter et al., 2012) using somatic copy-number. The cancer proportion estimated by the two methods showed a high correlation ( $\text{Cor} = 0.81$ ,  $p < 0.0001$ , Supplementary Figure S1C). Then we estimated the cell type gene expression through a constrained least squares fit (Onuchic et al., 2016) based on the cell type proportions obtained by deconvolution.

## Statistical analysis

All the statistical analyses were performed under R software version 4.1.0. Cox regression and survival analysis were performed using the R package “survival”. Cutoffs were identified using the R package “survminer”. Differentially expressed genes were identified using the R package “edgeR”. Gene set enrichment analysis was performed using the online tool Metascape (Zhou et al., 2019).

## Results

### Identification of six squamous malignant patterns across multiple cancer types

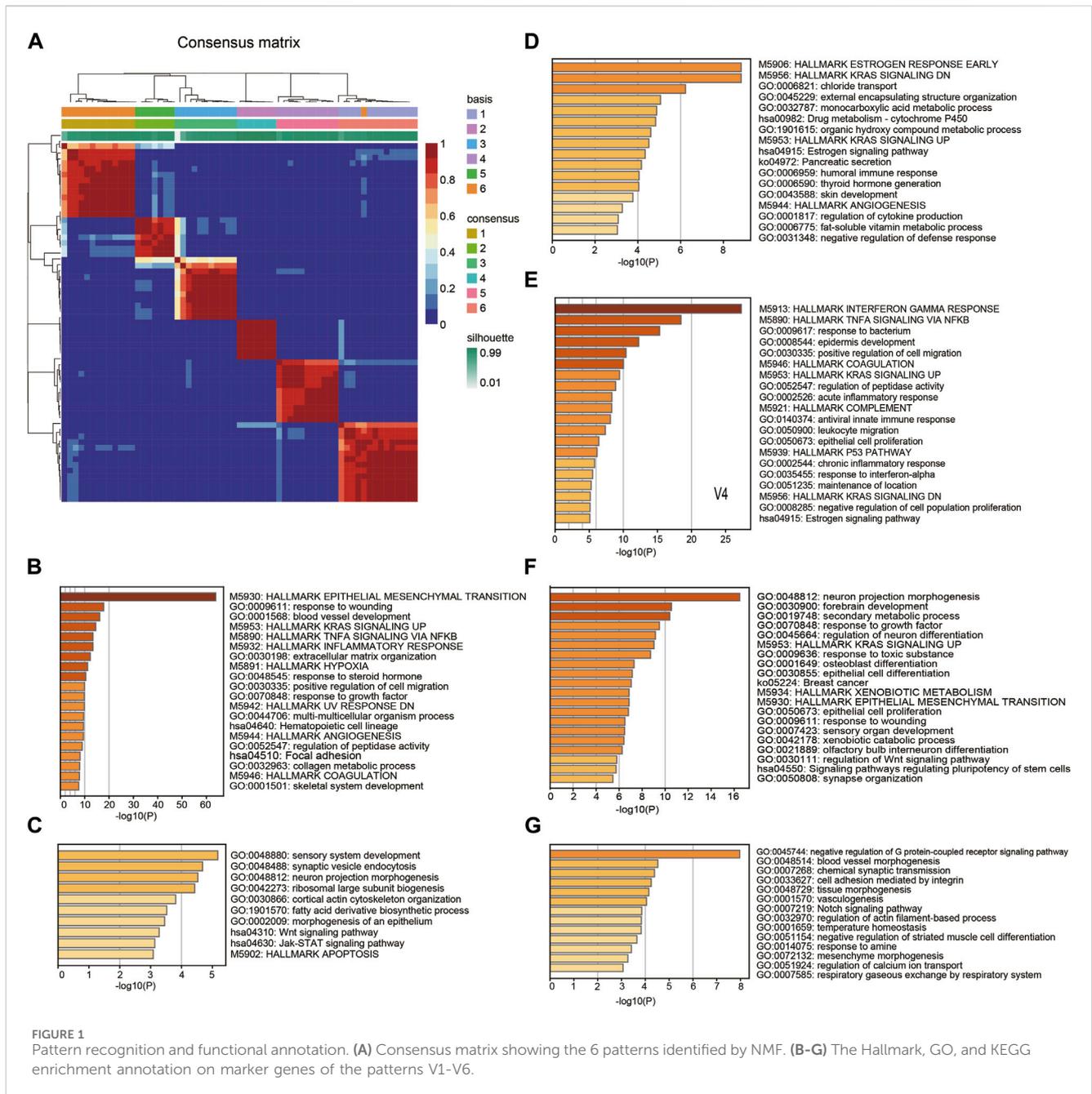
To explore the malignant gene patterns of squamous cancer cells, we studied 63 squamous cancer cell lines from a Sanger cell line Affymetrix gene expression project (GSE68950). These squamous cancer cell lines are derived from different parts of the body, including head and neck, lung, esophagus, cervix, and breast (Supplementary Table S2). We applied nonnegative matrix factorization (NMF) (Brunet et al., 2004) to identify molecular patterns of the cell line gene expression matrix. The algorithm recognized six patterns (V1-V6) across 63 cell lines, with the consensus matrix shown in Figure 1 A. We used Kim and Park’s gene scoring schema (Kim and Park, 2007) to extract the most relevant genes for each pattern (Supplementary Table S3). Enrichment analysis of these marker genes reflects the biological function of the pattern they belong to. As shown in Figure 1B, V1 contains many epithelial-mesenchymal transition (EMT) markers and it is also linked to other terms known to facilitate cancer progression, such as inflammatory response, hypoxia, positive regulation of cell migration, and angiogenesis. V2 is mainly associated with several growth and synthesis functions (Figure 1C); V3 is mainly related to metabolic functions (Figure 1D); V4 reflects cancer cell response to IFN- $\gamma$  and TNF- $\alpha$  (Figure 1E); V5 seems to participate in multiple morphogenesis and cell differentiation processes (Figure 1F); and V6 restricts G protein-coupled receptor signaling pathway and regulates cell adhesion (Figure 1G). It is worth noting that the enrichment analysis was based on prior curated gene sets with known functions (KEGG, GO and HALLMARK), while the NMF patterns were identified using matrix factorization without reference. Therefore, the function of a pattern is likely to be more complicated than a GO or KEGG term, and it is difficult to summarize a pattern using one known term.

### Stromal expressed malignant patterns contribute to HNSCC subtype classification

Although the six squamous malignant patterns were identified in cancer cell lines, they may be shared by other non-malignant cells in TME. The cancer-associated stromal cells in TME share multiple features with the cancer cells in solid tumors, many of which have been proven to affect prognosis. To illustrate the distribution of the 6 NMF patterns among different cell types in squamous tumors, we projected the patterns to a single-cell RNA sequencing (scRNA-seq) dataset (Puram et al., 2017) using transfer learning (Sharma et al., 2020). The scRNA-seq dataset includes 5,902 cells from 18 head and neck squamous (HNSCC) tumors after initial quality controls. We selected 1,136 cancer cells and 113 stromal cells with all six projection  $p$  values less than 0.05. As shown in Figure 2A, patterns V2, V3, and V4 are mainly expressed by cancer cells, while V1 and V5 are mainly expressed by stromal cells. Although the distribution of V6 is more inclined to stromal cells, the difference is not as big as the other five patterns showed.

Puram et al. have previously identified six meta-signatures (partial EMT, Epithelial differentiation one and 2, Stress, Hypoxia, and Cell cycle) that reflect common expression programs of malignant cells in this HNSCC scRNA-seq dataset (Puram et al., 2017). In addition, the TCGA HNSCC subtypes (Cancer Genome Atlas Network, 2015) (Atypical, Mesenchymal, Basal, and Classical) derived from bulk tumor sequencing were also projected to this single-cell dataset. It is worth noting that the Mesenchymal subtype has been validated as a non-malignant cell contributed signature in bulk tumor sequencing, rather than an independent cancer cell subtype. The other three subtypes have been validated by scRNA-seq (Puram et al., 2017). We evaluated the correlation of V1-V6 with these two groups of signatures in the 1,136 cancer cells and 113 stromal cells (Figure 2B). Pattern V1 was correlated to the partial EMT (p-EMT) signature (Figure 2B), which was identified as an independent predictor of nodal metastasis, grade, and adverse pathologic features (Puram et al., 2017). The marker genes of pattern V1 such as SERPINE1, ITGA5, TNC, EMP3, VIM, FN1 and THBS2 are also key genes related to p-EMT. However, pattern V1 could not be caused by EMT process because it was expressed in mesenchymal cells. This is also the key difference between pattern V1 and the p-EMT signature expressed by cancer cells. Pattern V1 also showed a high correlation with the Mesenchymal subtype signature, which is consistent with the high expression in stromal cells mentioned in Figure 2A. Therefore, we would like to explain V1 as a stromal-dominated pattern that coincides with the expression of EMT markers (Vellinga et al., 2016).

However, the other stromal-dominated pattern V5 showed almost no correlation with the Mesenchymal subtype signature, but a strong correlation ( $\text{cor} = 0.70$ ) with the Classical subtype signature and a moderate correlation ( $\text{cor} = 0.45$ ) with the Atypical subtype signature. It indicated that the Mesenchymal subtype signature did not capture all the features expressed by the stroma, and the cancer-associated stromal cells also contributed to the Classical and Atypical subtype signal from bulk tissue sequencing. As shown in Figure 2C, the stromal cells expressed an extremely low Basal signal but a relatively higher



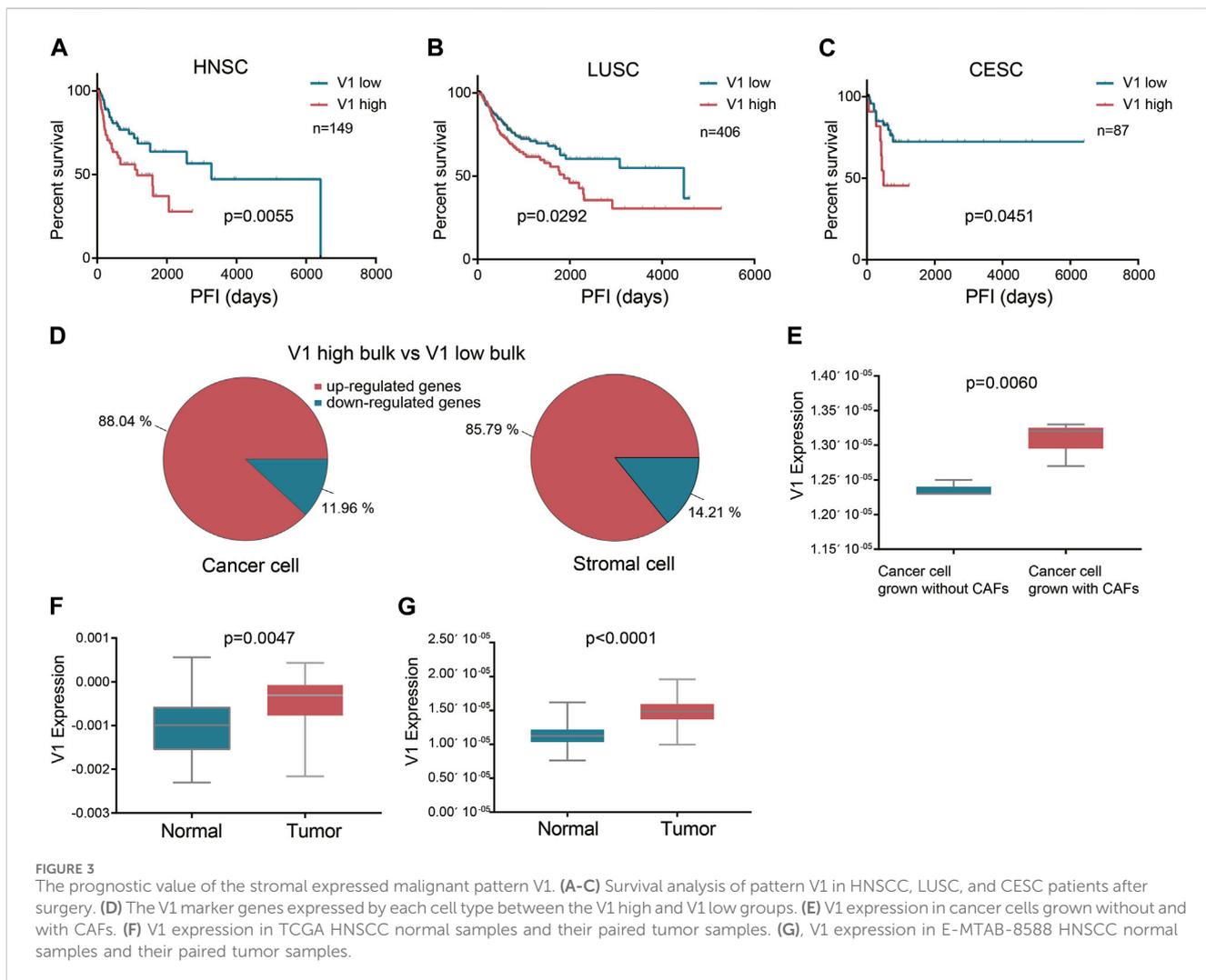
Atypical and Classical signal. The expression differences of Classical and Atypical signatures between the cancer cells vs. stromal cells are less than those of Basal and Mesenchymal subtypes. The correlation between Classical (Atypical) subtype signature and the stromal pattern V5 may explain this phenomenon. This finding validated that the Mesenchymal signature is derived from the stromal compartment and the Basal signature is derived from the cancer cells, while the Classical and Atypical signatures are mixed signals from both cancer cells and stromal cells.

As for cancer-dominated patterns, V2 and V4 were correlated to Basal subtype and Epithelial differentiation signature. Pattern V2 may contribute to the cell cycle signature while pattern V4 may participate in hypoxia. Pattern V3 only showed a moderate correlation with Atypical subtype.

## Stromal expressed malignant patterns are associated with poor PFI across different cancer types

To verify the influence of these squamous associated patterns on tumor progression and metastasis, we projected the six patterns to TCGA HNSCC, lung squamous cell carcinoma (LUSC), and cervical squamous cell carcinoma (CESC) datasets using transfer learning. Curated progression-free interval (PFI) recommended by the Pan-Cancer TCGA Atlas (Liu et al., 2018) was selected as the endpoint. In all the three TCGA cohorts, most patients received surgery alone before the first PFI event. In HNSCC cohort, surgery with radiotherapy was also common, while in CESC cohort, nearly half of the patients received surgery with radiochemotherapy. We found that high



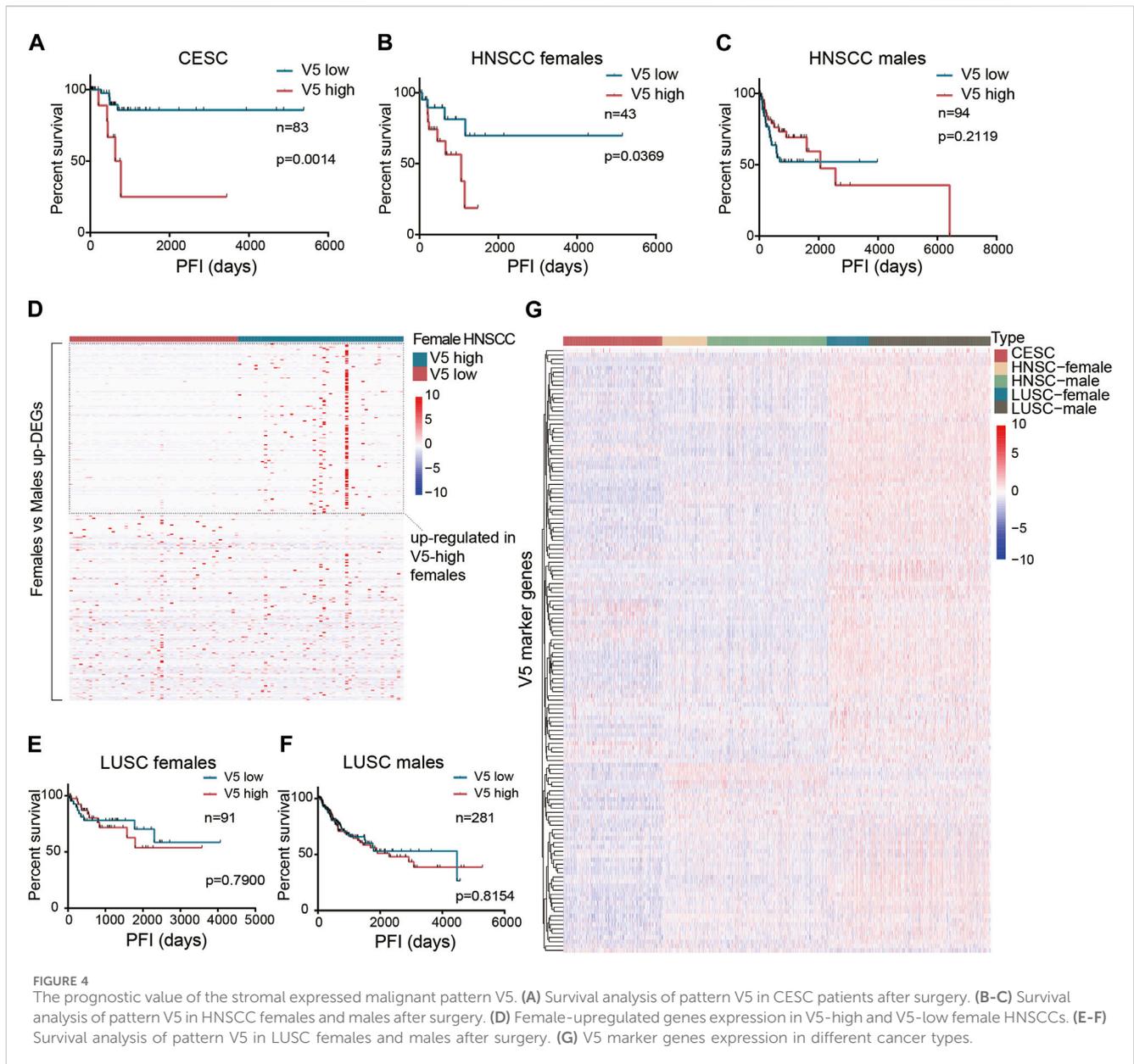


expressed lower V1 compared to their paired tumor ( $p = 0.0047$ , Figure 3F). We observed the same phenomenon in another cohort (E-MTAB-8588) with 98 paired normal and HNSCC tumor samples ( $p < 0.0001$ , Figure 3G).

In the TCGA LUSC cohort, most patients received surgery alone before the PFI event, while in the HNSCC cohort, surgery with radiotherapy is quite common. We selected 80 HNSCC patients who received surgery (margin negative) and full course adjuvant radiotherapy before the PFI event to perform the univariate cox analysis. The result showed that V1 expression had no influence on the PFI after surgery plus radiotherapy ( $p = 0.3073$ ). However, when we used a well-validated radiosensitivity score RSI (Eschrich et al., 2009) to divide those patients into radiosensitive (RS) and radioresistant (RR) groups, we found in the RS group V1 expression did not affect the PFI, while in the RR group high V1 was related to worse PFI after surgery plus radiotherapy, although with a  $p$ -value of 0.0532. Therefore we assume that SCCs with high V1 expression show a high tendency to recurrence and metastasis even after surgery with a negative margin, but radiation delivered to the tumor bed and high-risk areas after surgery may weaken the effect of high V1, depending on the tumor radiosensitivity. We did not further investigate the effect of V1 on PFI after chemotherapy because there was large heterogeneity in the

application of chemotherapy, and in most cases, chemotherapy was implemented with radiotherapy simultaneously.

Unlike the general prognostic value of pattern V1, we found that pattern V5 affected the PFI after surgery only in the CESC dataset (Figure 4A). Since CESC is special because it is female-only, we continued to study whether gender limits the prognostic value of pattern V5. After setting gender subgroups, we re-examined the effect of V5 on PFI after surgery (median V5 was used as cutoff). Interestingly, higher V5 did show a worse PFI in HNSCC females ( $p = 0.0369$ , Figure 4B) rather than in males (Figure 4C). In HNSCC males, the curves even showed an opposite tendency. The reason why V5 showed such a different effect on different sex was not clear. But we did identify 315 upregulated genes in female HNSCCs compared with male HNSCCs ( $FDR < 0.05$ , fold change  $> 2$ ), and half of them (152 in 315) were also upregulated in high V5 females compared with low V5 females (Figure 4D), which partially explains the gender bias of high-V5 effect. However, in the LUSC dataset, we did not observe a significant difference (Figures 4E,F): the survival curves of high V5 and low V5 showed a crossover in females. As shown in Figure 4G, the overall expression of V5 marker genes in LUSC is quite different from that in HNSCC and CESC, indicating an expression-dependent cancer-type specificity of V5's prognostic ability.



**FIGURE 4** The prognostic value of the stromal expressed malignant pattern V5. (A) Survival analysis of pattern V5 in CESC patients after surgery. (B-C) Survival analysis of pattern V5 in HNSCC females and males after surgery. (D) Female-upregulated genes expression in V5-high and V5-low female HNSCCs. (E-F) Survival analysis of pattern V5 in LUSC females and males after surgery. (G) V5 marker genes expression in different cancer types.

## Discussion

The study aimed to explore the malignant role of stroma in SCCs across different sites. We choose NMF for pattern recognition not only because of its excellent performance on the transcriptome but also because it considers all genes in the transcriptome without any reference or functional annotation. The algorithm always finds the most significant patterns and makes sure the combination of the patterns accurately reflects the target. That explains why the NMF algorithm often identifies cell type features when it is run in bulk sequencing data, since cell type usually contributes to the most significant differences within tumor tissue. To get the gene patterns reflecting the biological characteristics at the cell level, we performed NMF in sequencing data of the same cell type (cell line RNA-seq). Since NMF may not recognize malignant patterns from numerous non-malignant

features in stromal cells and cancer stroma shares many genetic features with cancer cells, we performed NMF in squamous cancer cell lines and proved that two malignant patterns are mainly expressed in cancer stroma.

The four-subtype classification of HNSCC based on bulk analyses (Atypical, Mesenchymal, Basal, and Classical) had been widely accepted until Puram et al. proved that no malignant cells in a scRNA-seq dataset mapped to the Mesenchymal subtype. They found the Mesenchymal signature in bulk sequencing was actually a stromal signal coming from the Basal subtype tumors. They also validated that the malignant cells of each tumor did map exclusively to one of the other three subtypes (Atypical, Basal, and Classical), even after control for TME. However, there have been no signatures describing the stromal composition of the Atypical and Classical bulk subtypes. In this study, we found a stroma-dominated pattern V5 showing a high correlation with the Classical signature,

moderate correlation with the Atypical signature, but no correlation with the Mesenchymal signature (Figure 2B). This pattern reflected a new malignant feature expressed by stromal cells of Classical and Atypical HNSCC tumors. Our results validated that the HNSCC Mesenchymal signature is derived from the stromal compartment and the Basal signature is derived from the cancer cells, while the Classical and Atypical signatures are mixed signals from both cancer and stromal cells. In TCGA CESC and HNSCC females, high expression of pattern V5 was associated with poor PFI. Further study may focus on the gender-biased prognostic mechanism of pattern V5.

Another stroma-dominated pattern V1 is associated with poor PFI across different cancer types (HNSCC, LUSC, and CESC). The marker genes of pattern V1 were enriched in the Hallmark EMT gene set. However, we believe this change is unrelated to EMT because V1 was mainly expressed by mesenchymal cells. In fact, researches on colorectal and urothelial cancers have shown that stromal cells are a major source of EMT-related gene expression in bulk transcriptomes, rather than epithelial-derived cancer cells (Isella et al., 2015; Wang et al., 2018; Li et al., 2023; Cañellas-Socias et al., 2024), which seems contrary to the biological concept of EMT. This phenomenon could be explained by an activated stromal state that coincides with cancer cell expressed EMT-related markers (Vellinga et al., 2016). Although these genes are clearly expressed in cancer cells after EMT, the expression level was far lower than that in CAFs. In addition, we found that V1 expression in cancer cells could be induced by co-culturing with CAFs. A similar phenomenon has been found in colorectal cancer that cancer cells express mesenchymal genes in a high-stroma context (Vellinga et al., 2016). Pattern V1 demonstrates the key effect of cancer stroma on the progression and metastasis of different types of squamous cell carcinoma. Interestingly, radiation delivered to the tumor bed and high-risk areas after surgery may weaken the effect of high V1 on recurrence and metastasis, depending on the tumor radiosensitivity. Further study may compare the different outcomes of surgery alone and surgery plus radiotherapy in radiosensitive tumors with high and low V1 expression, respectively. The pattern V1 may bring a new indication for postoperative radiotherapy.

Previous theories have been considering cancer stroma as “dangerous liaisons” interacting with cancer cells and other components in TME to form a cancer-supportive environment (Chen and Song, 2019; Gieniec et al., 2019). In this study, we confirmed the prognostic value of stromal gene features and its universality in SCCs. Therefore, We suggest that the tumor genetic subtype classification may be changed to a new system that integrates both malignant and non-malignant components. There are potential limitations to this work, given the lack of validation in cell experiments. Adding this part of the experimental content in the future may provide evidence for the application of this research in the clinical field.

## Conclusion

In this study, we confirmed the prognostic value of stromal gene features and its universality. Therefore, we suggest that the tumor genetic subtype classification may be changed to a new system that integrates both malignant and non-malignant components.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

## Author contributions

KQ: Writing—original draft. GL: Writing—original draft. YJ: Writing—review and editing. XT: Writing—review and editing. QQ: Writing—review and editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the National Natural Science Foundation of China (Grant No. 82272749) and Wu Jieping Medical Foundation (No. 320.6750.2022-17-54).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2024.1342306/full#supplementary-material>

### SUPPLEMENTARY TABLE S1

A total of 303 GEO DNA methylation profiles of known cell types as reference for deconvolution.

### SUPPLEMENTARY TABLE S2

Squamous cancer cell lines from the Sanger cell line Affymetrix gene expression project (GSE68950). These squamous cancer cell lines are derived from different parts of the body, including head and neck, lung, esophagus, cervix, and breast.

### SUPPLEMENTARY TABLE S3

Marker genes of each pattern identified by NMF.

### SUPPLEMENTARY FIGURE S1

Deconvolution of the TCGA HNSCC dataset.

### SUPPLEMENTARY FILE S1

Details of data analysis.

## References

- Baumeister, P., Zhou, J., Canis, M., and Gires, O. (2021). Epithelial-to-Mesenchymal transition-derived heterogeneity in head and neck squamous cell carcinomas. *Cancers (Basel)* 13 (21), 5355. doi:10.3390/cancers13215355
- Bhowmick, N. A., Neilson, E. G., and Moses, H. L. (2004). Stromal fibroblasts in cancer initiation and progression. *Nature* 432 (7015), 332–337. doi:10.1038/nature03096
- Brunet, J. P., Tamayo, P., Golub, T. R., and Mesirov, J. P. (2004). Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. U. S. A.* 101 (12), 4164–4169. doi:10.1073/pnas.0308531101
- Cancer Genome Atlas Network (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* 517 (7536), 576–582. doi:10.1038/nature14129
- Cañellas-Socias, A., Sancho, E., and Batlle, E. (2024). Mechanisms of metastatic colorectal cancer. *Nat. Rev. Gastroenterol. Hepatol.* doi:10.1038/s41575-024-00934-z
- Carter, S. L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* 30 (5), 413–421. doi:10.1038/nbt.2203
- Chai, A. W. Y., Lim, K. P., and Cheong, S. C. (2020). Translational genomics and recent advances in oral squamous cell carcinoma. *Semin. Cancer Biol.* 61, 71–83. doi:10.1016/j.semcancer.2019.09.011
- Chen, X., and Song, E. (2019). Turning foes to friends: targeting cancer-associated fibroblasts. *Nat. Rev. Drug Discov.* 18 (2), 99–115. doi:10.1038/s41573-018-0004-1
- Chhabra, Y., and Weeraratna, A. T. (2023). Fibroblasts in cancer: unity in heterogeneity. *Cell* 186 (8), 1580–1609. doi:10.1016/j.cell.2023.03.016
- Cox, T. R. (2021). The matrix in cancer. *Nat. Rev. Cancer* 21 (4), 217–238. doi:10.1038/s41568-020-00329-7
- Devarajan, K. (2008). Nonnegative matrix factorization: an analytical and interpretive tool in computational biology. *PLoS Comput. Biol.* 4 (7), e1000029. doi:10.1371/journal.pcbi.1000029
- Dotto, G. P. (2014). Multifocal epithelial tumors and field cancerization: stroma as a primary determinant. *J. Clin. Invest.* 124 (4), 1446–1453. doi:10.1172/JCI72589
- Eschrich, S. A., Pramana, J., Zhang, H., Zhao, H., Boulware, D., Lee, J. H., et al. (2009). A gene expression model of intrinsic tumor radiosensitivity: prediction of response and prognosis after chemoradiation. *Int. J. Radiat. Oncol. Biol. Phys.* 75 (2), 489–496. doi:10.1016/j.ijrobp.2009.06.014
- Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinforma.* 11, 367. doi:10.1186/1471-2105-11-367
- Gieniec, K. A., Butler, L. M., Worthley, D. L., and Woods, S. L. (2019). Cancer-associated fibroblasts—heroes or villains? *Br. J. Cancer* 121 (4), 293–302. doi:10.1038/s41416-019-0509-3
- Guan, Y., Wang, G., Fails, D., Nagarajan, P., and Ge, Y. (2020). Unraveling cancer lineage drivers in squamous cell carcinomas. *Pharmacol. Ther.* 206, 107448. doi:10.1016/j.pharmthera.2019.107448
- Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell* 144 (5), 646–674. doi:10.1016/j.cell.2011.02.013
- Isella, C., Terrasi, A., Bellomo, S. E., Petti, C., Galatola, G., Muratore, A., et al. (2015). Stromal contribution to the colorectal cancer transcriptome. *Nat. Genet.* 47 (4), 312–319. doi:10.1038/ng.3224
- Kim, H., and Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics* 23 (12), 1495–1502. doi:10.1093/bioinformatics/btm134
- Li, G., Jiang, Y., Li, G., and Qiao, Q. (2021). Comprehensive analysis of radiosensitivity in head and neck squamous cell carcinoma. *Radiother. Oncol.* 159, 126–135. doi:10.1016/j.radonc.2021.03.017
- Li, H., Courtois, E. T., Sengupta, D., Tan, Y., Chen, K. H., Goh, J. J. L., et al. (2023). Publisher Correction: reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.* 55 (1), 166. doi:10.1038/s41588-022-01281-y
- Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., et al. (2018). An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* 173 (2), 400–416.e11. doi:10.1016/j.cell.2018.02.052
- NCHBI (2021). Expression data of HNSCC cell lines grown without and with patient-matched CAFs in 2D. Available at: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE178153>.
- NCHBI (2015). Sanger cell line Affymetrix gene expression project. Available at: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE68950>.
- NCHBI (2023). Gene expression Omnibus. Available at: <https://www.ncbi.nlm.nih.gov/geo/>.
- National Cancer Institute (2024). Genomic data commons data portal. Available at: <https://portal.gdc.cancer.gov/>.
- Onuchic, V., Hartmaier, R. J., Boone, D. N., Samuels, M. L., Patel, R. Y., White, W. M., et al. (2016). Epigenomic deconvolution of breast tumors reveals metabolic coupling between constituent cell types. *Cell Rep.* 17 (8), 2075–2086. doi:10.1016/j.celrep.2016.10.057
- Plava, J., Cihova, M., Burikova, M., Matuskova, M., Kucerova, L., and Miklikova, S. (2019). Recent advances in understanding tumor stroma-mediated chemoresistance in breast cancer. *Mol. Cancer* 18 (1), 67. doi:10.1186/s12943-019-0960-z
- Puram, S. V., Tirosh, I., Park, A. S., Patel, A. P., Yizhak, K., Gillespie, S., et al. (2017). Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* 171 (7), 1611–1624. doi:10.1016/j.cell.2017.10.044
- Rahal, Z., Sinjab, A., Wistuba, I. I., and Kadara, H. (2022). Game of clones: battles in the field of carcinogenesis. *Pharmacol. Ther.* 237, 108251. doi:10.1016/j.pharmthera.2022.108251
- Schwaederle, M., Elkin, S. K., Tomson, B. N., Carter, J. L., and Kurzrock, R. (2015). Squamousness: next-generation sequencing reveals shared molecular features across squamous tumor types. *Cell Cycle* 14 (14), 2355–2361. doi:10.1080/15384101.2015.1053669
- Sharma, G., Colantuoni, C., Goff, L. A., Fertig, E. J., and Stein-O'Brien, G. (2020). projectR: an R/Bioconductor package for transfer learning via PCA, NMF, correlation and clustering. *Bioinformatics* 36 (11), 3592–3593. doi:10.1093/bioinformatics/btaa183
- Slaughter, D. P., Southwick, H. W., and Smejkal, W. (1953). Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer* 6 (5), 963–968. doi:10.1002/1097-0142(195309)6:5<963::aid-cnrc2820060515>3.0.co;2-q
- Smetana, K., Strnad, H., Kolar, M., Plzak, J., Novotny, J., and Chovanec, M. (2020). Transcription profiling by array of paired head and neck squamous cell carcinoma and normal mucosa samples. Available at: <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-8588/>.
- Valkenburg, K. C., de Groot, A. E., and Pienta, K. J. (2018). Targeting the tumour stroma to improve cancer therapy. *Nat. Rev. Clin. Oncol.* 15 (6), 366–381. doi:10.1038/s41571-018-0007-1
- Vellinga, T. T., den Uil, S., Rinkes, I. H. B., Marvin, D., Ponsioen, B., Alvarez-Varela, A., et al. (2016). Collagen-rich stroma in aggressive colon tumors induces mesenchymal gene expression and tumor cell invasion. *Oncogene* 35 (40), 5263–5271. doi:10.1038/onc.2016.60
- Wang, L., Saci, A., Szabo, P. M., Chasalow, S. D., Castillo-Martin, M., Domingo-Domenech, J., et al. (2018). EMT- and stroma-related gene expression and resistance to PD-1 blockade in urothelial cancer. *Nat. Commun.* 9 (1), 3503. doi:10.1038/s41467-018-05992-x
- Xu, M., Zhang, T., Xia, R., Wei, Y., and Wei, X. (2022). Targeting the tumor stroma for cancer therapy. *Mol. Cancer* 21 (1), 208. doi:10.1186/s12943-022-01670-1
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10 (1), 1523. doi:10.1038/s41467-019-09234-6