Observational Study

Medicine

OPEN

Expression and Functional Analysis of core stemness factors OSKM (OCT4, SOX2, KLF4, and MYC) in Pan-cancer

Liwei Hong, PhD[a,b,c,*], Sijie Hong, MM[d], Xueqin Zhang, MM[a,c,e]

Abstract

The dedifferentiation process of tumorigenesis and somatic cell reprogramming has some commonness and differences, which is the key question to cancer therapeutic strategy and stem cell applications. To further explore the commonalities and variance between carcinogenesis and induced pluripotent stem cell reprogramming, we investigated the role of stemness factors OSKM (OCT4, SOX2, KLF4, and MYC) in the pan-cancer process using public clinical data. Expression of OSKM in human pan-cancer was analyzed via the Genotype Tissue Expression (GTEx) and The Cancer Genome Atlas (TCGA) database based on the RNA-seq data of tissues. The correlation of expression between OSKM genes was analyzed via the Tumor Immune Evaluation Resource (TIMER) database, while the STRING tool was used to construct the protein-protein interaction network for OSKM. Prognostic impact of OSKM in pan-cancer was analyzed by Cox proportional hazards regression model. The relationships between OSKM and tumor stemness, tumor microenvironment and immune checkpoint and were performed by Sangerbox platform using Pearson correlation analysis. Our results showed that OSKM were universally expressed and significantly altered in tumors compared with adjacent normal tissues in most tumor types. In addition, correlation analysis revealed the relevance of OSKM genes to patient prognosis, cancer cell stemness, tumor microenvironment or immune checkpoint. However, there is little similarity between these genes in terms of how they function in each cancer type. This study elucidates the different roles of core stemness factors OSKM in pan-cancer, offering potential therapeutic targets for novel anti-cancer strategies and knowledge to minimize the potential carcinogenic effects during stem cell transplantation.

Abbreviations: ACC = adrenocortical carcinoma, BLCA = Bladder Urothelial Carcinoma, BRCA = breast invasive carcinoma, CESC = cervical squamous cell carcinoma and endocervical adenocarcinoma, CHOL = cholangiocarcinoma, COAD = colon adenocarcinoma, COADREAD = colon adenocarcinoma/rectum adenocarcinoma esophageal carcinoma, DLBC = Lymphoid Neoplasm Diffuse Large B-cell Lymphoma, ESCA = esophageal carcinoma, GBM = glioblastoma multiforme, GBMLGG = glioma, HNSC = Head and Neck squamous cell carcinoma, KICH = Kidney Chromophobe, KIPAN = pan-kidney cohort (KICH+KIRC+KIRP), KIRC = kidney renal clear cell carcinoma, KIRP = kidney renal papillary cell carcinoma, KLF4 = Kruppel-like factor 4, LAML = Acute Myeloid Leukemia, LGG = Brain Lower Grade Glioma, LIHC = liver hepatocellular carcinoma, LUAD = lung adenocarcinoma, LUSC = lung squamous cell carcinoma, MESO = Mesothelioma, MYC = c-myelocytomatosis, NB = Neuroblastoma, OSKM = POU5F1 (OCT4), SOX2, KLF4 and MYC, OV = ovarian serous cystadenocarcinoma, PAAD = pancreatic adenocarcinoma, PCPG = pheochromocytoma and Paraganglioma, POU5F1 (OCT4) = octamer ¾, PRAD = prostate adenocarcinoma, READ = rectum adenocarcinoma, SARC = Sarcoma, SKCM = Skin Cutaneous Melanoma, SOX2 = SRY-box-containing gene 2, STAD = stomach adenocarcinoma, STES = stomach and esophageal carcinoma, TGCT = Testicular Germ Cell Tumors, THCA = thyroid carcinoma, THYM = thymoma, UCEC = Uterine Corpus Endometrial Carcinoma, UCS = uterine carcinosarcoma, UVM = uveal melanoma, WT = high-risk Wilms tumor.

Keywords: core pluripotency factors, immune checkpoint, pan-cancer, prognosis, tumor purity

This research was supported by the Xiamen Science and Technology Bureau to Xueqin Zhang (grant no. 3502Z20191102).

The authors have no conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are publicly available.

Supplemental Digital Content is available for this article.

[a] Xiamen Key Laboratory of Basic and Clinical Research on Major Obstetrical Diseases, Women and Children's Hospital, School of Medicine, Xiamen University, Xiamen, China, [b] Medical College, Nanchang Institute of Technology, Nanchang, China, [c] Xiamen Clinical Research Center for Perinatal Medicine, Women and Children's Hospital, School of Medicine, Xiamen University, Xiamen, China, [d] Department of Ultrasound, Women and Children's Hospital, School of Medicine, Xiamen University, Xiamen, China, [e] Department of Obstetrics, Women and Children's Hospital, School of Medicine, Xiamen University, Xiamen, China.

**Correspondence: Liwei Hong, Women and Children's Hospital, 10 Zhenhai Road, Xiamen, Fujian, China. e-mail: honglw@xmu.edu.cn*

How to cite this article: Hong L, Hong S, Zhang X. Expression and Functional Analysis of core stemness factors OSKM (OCT4, SOX2, KLF4, and MYC) in Pan-cancer. Medicine 2023;102:48(e36433).

Received: 11 September 2023 / Received in final form: 30 October 2023 / Accepted: 10 November 2023

http://dx.doi.org/10.1097/MD.0000000000036433

1. Introduction

Cancer, as a heavy burden for humans across the world, has become a serious threat endangering human life. The number of deaths caused by malignant tumors in the world exceeds 10 million each year. Of these, the mortality rate is high among all diseases.[1]

Although tumorigenesis is related to many reasons, it is believed that tumor is a stem cell disease,[2–4] which originates from the malignant transformation of tissue adult stem cells. Research has demonstrated that tumor cells exhibit distinct traits and behaviors that closely resemble stem cell characteristics, such as the ability to self-renew, unlimitedly proliferate and differentiate into various types of cells.[5] There are also high similarities between the dedifferentiation process of tumorigenesis and somatic cell reprogramming, suggesting that reprogramming factors of induced pluripotent stem cells (iPSC) might play important roles in oncogenesis.[6,7] Uncovering shared features with iPSC can unveil potential mechanisms underlying tumor initiation, progression, and metastasis.[8] The differences between tumorigenesis dedifferentiation and the reprogramming of somatic cells must be taken into account for the cancer therapy and application of iPSC, which helps interpret the unique characteristics driving cancer development, advancing our understanding of cancer biology. The differentiation between tumors and iPSCs also aids in identifying biomarkers for cancer diagnosis and prognosis, facilitating the development of more accurate and personalized diagnostic tools.[9,10] Moreover, it assists in better discrimination during stem cell therapies, ensuring the safety and efficacy of interventions based on stem cells and minimizing potential risks associated with cancer transformation.[11–13]

iPSCs was initially created from differentiated fibroblasts by Takahashi and Yamanaka via the overexpression of 4 specific genes OSKM (octamer 3/4 (OCT4 or POU5F1), SRY-box-containing gene 2 (SOX2), Kruppel-like factor 4 (KLF4) and c-myelocytomatosis (MYC)).[14] This revolutionary technology in human cells was eventually accomplished 1 year later by researchers successively.[15] It was also demonstrated that lin28 homolog A (LIN28A) and Nanog homeobox (NANOG) can replace KLF4 and MYC in the Yamanaka reprogramming cocktail.[16,17] Since then, iPSC were well studied and currently being evaluated in regenerative therapy proceeding towards clinical trials.[18–20] Considering about 20% of mice offspring derived from iPSCs were found to develop tumors, it was questioned that the attributes of so-called stem cell is more like those of tumor cells and MYC, KLF4 genes might be carcinogenic.[21] Even though more and more studies indicated that the stem cell transcription factors play important roles in maintaining the self-renewal, reprogramming and multiple differentiations of tumor cells,[22,23] how OSKM genes participate in these processes still needs to be further explored.

With the continuous development and improvement of public databases, pan-cancer analysis uncovers the commonness and differences of tumors and provides valuable insights into cancer prevention and candidate therapeutic targets design,[24] which might offer a new way to further explore the commonalities and variance between carcinogenesis and induced pluripotent stem cell reprogramming. In this study, we systematically examined the transcription and translation of key reprogramming factors OSKM in paired normal and cancer tissues by combining data from various databases. Then the expression association and protein-protein interaction network between OSKM genes was accessed. Meanwhile, we evaluated the prognostic value of OSKM in pan-cancer based on The Cancer Genome Atlas (TCGA) dataset. Subsequently, we explored the correlation between the OSKM expression levels and tumor stemness, tumor microenvironment and immune checkpoints. Our pan-cancer analysis provides insight into the role of core stemness factors OSKM in the development of pan-cancer, and demonstrate that OSKM-mediated mechanisms of cancer regulation are distinct from their synergistic regulation of iPSC reprogramming as previously reported.

2. Methods

2.1. Data source

The Genotype-Tissue Expression (GTEx) (https://www.gtexportal.org) established a data resource and tissue bank to study the relationship between genetic variation and gene expression in multiple human tissues.[25] The Cancer Genome Atlas Program (TCGA) (https://www.cancer.gov/ccg/research/genome-sequencing/tcga) is a free data portal of landmark cancer genomics program, providing clinic and pathological information of 33 types of cancer for researchers.[26] The expression of OSKM in normal and tumor tissues was accessed and analyzed by combining the data from the GTEx and TCGA database via a clinical bioinformatics analysis platform Sangerbox (http://vip.sangerbox.com/).[27] Samples with undetected expression levels were filtered, and cancer species with less than 3 samples in a single cancer specie were also excluded. All expression data were normalized through log2(x + 0.001) conversion.

2.2. Gene correlation analysis

The Tumor Immune Estimation Resource (TIMER) is a comprehensive resource for systematical analysis of immune infiltrates across diverse cancer types (http://timer.cistrome.org/),[28] in which "Gene_Corr" module allows users to discover the co-expression pattern of genes inputted across TCGA cancer types using Pearson's correlation analysis.[29] The correlation between the mRNA expression of OSKM or with 60 marker genes related immune checkpoint pathway (24 inhibitory and 36 stimulatory genes) was evaluated with the purity-adjusted partial spearman's rho value as the degree of their correlation. The correlation between OSKM genes and target signature across all TCGA tumors was drawn by R language with "heatmap" package.

2.3. Protein-protein interaction comprehensive analysis

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (https://string-db.org/) is a precomputed global resource for the exploration and analysis of functional links between proteins, which currently covers 67'592'464 proteins from 14'094 organisms and has been proven effective in extensive research.[30] STRING has a unique scoring system that integrates a single confidence score for each prediction with benchmarks of the various kinds of associations against a common reference set. After importing the OSKM into the online tool STRING version 11.5, we obtained the protein–protein interaction network information about the Yamanaka factors.

2.4. Prognosis analysis

To investigate the effects of OSKM expression levels on the clinical outcomes, we constructed a prognostic classifier using Kaplan–Meier (KM) survival curves to compare the survival disparities. Data was downloaded from UCSC (https://xenabrowser.net/) and a previously published TCGA prognostic study[31] to obtain a high-quality prognostic dataset of TCGA and eliminate samples with a follow-up time shorter than 30 days or cancers with less than 10 samples in a single cancer species. The coxph function of R software package survival (version 3.2-7) was used to establish the Cox proportional hazards regression model for analysis of the correlation between OSKM expression and prognosis in each tumor.[32] Statistical test was conducted by Logrank test to obtain the significance of prognosis. The correlation between OSKM

expression and patients' prognosis, such as overall survival (OS) and disease-specific survival (DSS), was shown by forest plots. The log-rank *P*-value and hazard ratio (HR) with 95% confidence intervals were calculated via univariate survival analysis.

2.5. Stemness features analysis of tumors

The correlations between OSKM genes and stemness features were analyzed by online tool Sangerbox (http://vip.sangerbox.com/) using Pearson's correlation method with all types of TCGA tumor samples. The cancer species with less than 3 samples in a single cancer species was eliminated in stemness features analysis, and we finally obtained the expression data of 37 cancer species. Stemness Scores were calculated using mRNA and DNA methylation profiles based on the previous study, including DNA methylation-based stemness scores (DNAss) and mRNA expression-based stemness scores (RNAss).[33]

2.6. Immune microenvironment of tumors

ESTIMATE algorithm (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data) is a tool to predict tumor purity and the presence of infiltrating atromal/immune cells in tumor tissues using gene expression data.[34] ESTIMATE algorithm is based on single sample Gene Set Enrichment Analysis and generates 3 scores (stromal score, immune score and ESTIMATE score). Scores of all tumor samples from the TCGA database were assessed by "estimate" R package in R 3.6.0 in this study. Then Spearman's correlation test was performed to examine the correlation between the expression of OSKM genes and these 3 scores.

2.7. Ethics statements

The patients with the gene expression data and matching clinical information were obtained by the UCSC cancer genomics browser, which is publicly open-access and therefore there is no need of the local ethics committee or informed consent.

3. Results

3.1. Expression of core reprogramming factors changed in pan-cancer

To investigate the potential involvement of those reprogramming factors in carcinogenesis and progression, we initially evaluated the mRNA expression of 6 core pluripotency factors POU5F1, SOX2, KLF4, MYC, LIN28A, and NANOG in 34 human tumors and corresponding adjacent normal tissues using a combination of data from the TCGA and GTEx database. The results of pan-cancer examination showed that transcription of OSKM was detected in all tissues as expected. And consistent with our assumptions, the expression of all 6 detected genes was changed in most of the cancer tissues compared with respective healthy tissues with significant difference, suggesting that 6 reprogramming factors play important roles in tumorigenesis. However, trends in expression of individual genes in cancer are different. For example, KLF4, MYC and NANOG were down-regulated in ovarian serous cystadenocarcinoma (OV) tumor tissues, but mRNA levels of POU5F1 and LIN28A were significantly higher, while there was no significant change for SOX2. Furthermore, the mRNA levels of LIN28A and NANOG were extremely low in almost all types of normal and cancer tissues except for TGCT. The findings were displayed in Figure 1, which were consistent with those calculated with the Gene expression profiling interactive analysis (GEPIA) online server Figure S1, Supplemental Digital Content, http://links.lww.com/MD/K925). To avoid the analysis error caused by low expression of LIN28A and NANOG, we removed them from the candidate list of subsequent analysis in this study. Taken together, the various transcription levels of 6 detected stemness genes in pan-cancer indicates a complex network for core reprogramming factors in regulation of cancers.

To further verify the changes of 4 key stemness factors at the protein level in tumors, we accessed the antibody-based IHC analyses across different cancer types via Human Protein Atlas (HPA). Corresponding to their transcription levels, the translation of OSKM genes also changed in various cancer tissues compared with paired normal tissues (Figure S2, Supplemental Digital Content, http://links.lww.com/MD/K926).

3.2. Relationship between OSKM in pan-cancer

Regulation of gene expression and protein interactions are important aspects that influence the functions they perform. In order to further examine the potential interaction of OSKM gene in the carcinogenesis process, we carried out the correlation analysis between OSKM genes in various cancer types by Gene_Corr module of Tumor Immune Estimation Resource (TIMER), a comprehensive resource for systematical evaluations of genes across diverse cancer types. The *P* value and the correlation coefficient (spearman's rho) were displayed in the heatmap (Fig. 2A). Our results showed that the expression of KLF4 was positively correlated with the others 3 genes in most types of cancer. Although also with significance, the correlation among OSM was more complex.

In addition to expression correlation, the functional interaction between proteins is necessary for the molecular mechanism of malignancy. Therefore, we used STRING tool to construct the protein-protein interaction network for OSKM (Fig. 2B). Each edge in the network denotes a known or predicted interaction. We found that OSKM proteins interact with each other. Also, important nodes that affect stem cell phenotype, immune escape, tumor growth and metastasis of cancers could be found in this network, such as SP1, EP300, CTNNB1, SKP2, FBXW7, MAX, and KAT family, implying a molecular function of OSKM in regulation of cancers.

3.3. Prognostic impact of OSKM in Pan-Cancer

Since expression of OSKM altered significantly in various tumors compared with normal tissues, we suppose that transcription of these genes significantly might associate with cancer prognostic. Survival is the most reliable primary end-point to assess the prognosis in oncologic clinical studies. Using the Cox proportional hazards regression model, we calculated correlation between the respective expression level of OSKM and overall survival (OS) in different cancer types, using data from TCGA database. The corresponding risk values are shown in forest plots (Fig. 3A). The findings demonstrated that OSKM mainly participated in the prognosis of BLCA, GBMLGG and kidney cancers (KIRP, KIRC, and KIPAN).

In order to avoid the major limitation of overall survival, the inclusion of non-cancer related deaths, disease-specific survival (DSS) is increasingly used as an advanced prognostic parameter and represents better the extent and reliability of prognostic evidence for patients with cancer. Therefore, the correlation between OSKM and DSS were also examined in the cancers mentioned above, except LAML due to lack of related data (Fig. 3B). Our findings showed that the results obtained by OS and DSS are elementarily agreeable.

3.4. Pan-cancer stemness landscape affected by OSKM

According to previous studies, the primary factors influencing the growth of tumors are the acquisition of stem cell-like features, such as self-renewal and dedifferentiation.[35] DNA

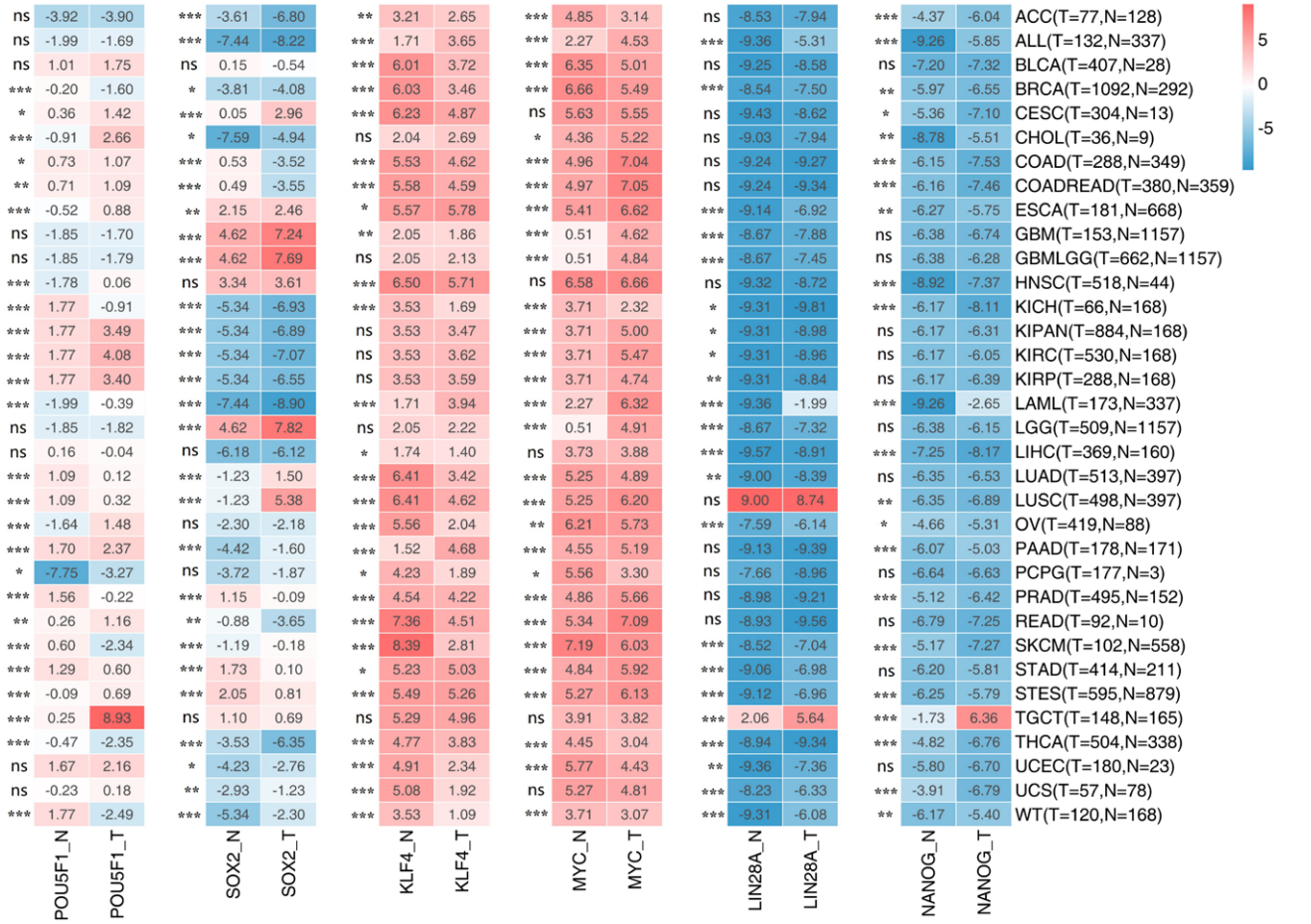

Figure 1. Differential mRNA expression of core reprogramming factors in paired normal tissues and tumors. The numbers represents mean of expression after log2(x + 0.001) conversion, N = normal tissues, T = tumor. Unpaired Wilcoxon Rank Sum and Signed Rank Tests were used to perform significance of difference analyses, ns = no significant difference, *$P < .05$, **$P < .01$, ***$P < .001$. OSKM = POU5F1 (OCT4), SOX2, KLF4 and MYC.

methylation-based Stemness Score (DNAss) and RNA expression-based Stemness Score (RNAss) measure the similarities between tumor cells and stem cells. A higher stemness score means that the tumor cells are more similar to stem cells. Considering OSKM genes were key reprogramming factors for stem cell pluripotency, we investigated the relationship between expression of OSKM and tumor stemness indices in 37 tumor types. Our results revealed that stemness of cancer cell in almost all tumor types was significantly affected by at least 1 gene among OSKM, suggesting that Yamanaka factors play roles in regulating cancer cell stemness (Fig. 4). However, consistency of each gene between DNAss and RNAss was poor, as were the stemness results of inter-gene comparisons, implying intricate entanglements of OSKM in regulating cancer cell stemness.

3.5. Relationship between OSKM expression and the tumor microenvironment

Together with cancer cells, the tumor microenvironment (TME) is made up of a mass of heterogeneous cell types, such as stromal cells and immune cells.[36] The tumor microenvironment contributes to the stemness maintenance, angiogenesis, invasion, metastasis and chronic inflammation of cancers. In order to account for OSKM's impact on the tumor microenvironment, we calculate the Stromal, Immune and Estimate scores in 37 types of cancer by using ESTIMATE algorithm to predict the level of infiltrating stromal, immune cells and total tumor purity respectively, and then evaluated whether a relevance exists between OSKM gene transcription levels and these 3 scores. Our results showed tumor purity of all types of cancers were affected by OSKM. OSKM expression had a significantly negative effect on 3 scores in HNSC, LUSC, and STES, suggesting that those key stemness factors produce a markedly positive effect on the tumor purity of these 3 types of cancers. In addition, we found that KLF4positivelycorrelated with Stromal, Immune and Estimate scores in the majority of cancer kinds, while SOX2played as negative regulator (Fig. 5), indicating different roles of OSKM in tumor microenvironment regulation.

3.6. Correlation of OSKM with immune checkpoint genes

The immune system recognizes and eliminates cancer, which regulated by an exquisite system of checks and balances. There are some key genes strongly associated with and recognized as checkpoint components in the immune response, which involve both co-stimulatory and inhibitory proteins.[37] Stimulatory checkpoint pathways can promote cell survival, cell cycle progression and effector/memory cells differentiation, whereas inhibitory checkpoint pathways terminate T cell activation to halt ongoing inflammation.[38] To further investigate the role of OSKM genes in immune regulation for cancers, the correlation of OSKM and immune checkpoint gene expression in various tumor types was calculated. The Pearson's correlation coefficient between OSKM expression and 24 inhibitory genes as well as 36 stimulatory genes was analyzed. Our results showed that almost all immune checkpoint genes were significantly positively co-expressed with KLF4 and MYC in most kinds of tumors except for LUSC (Fig. 6). Immune checkpoint genes were suppressed

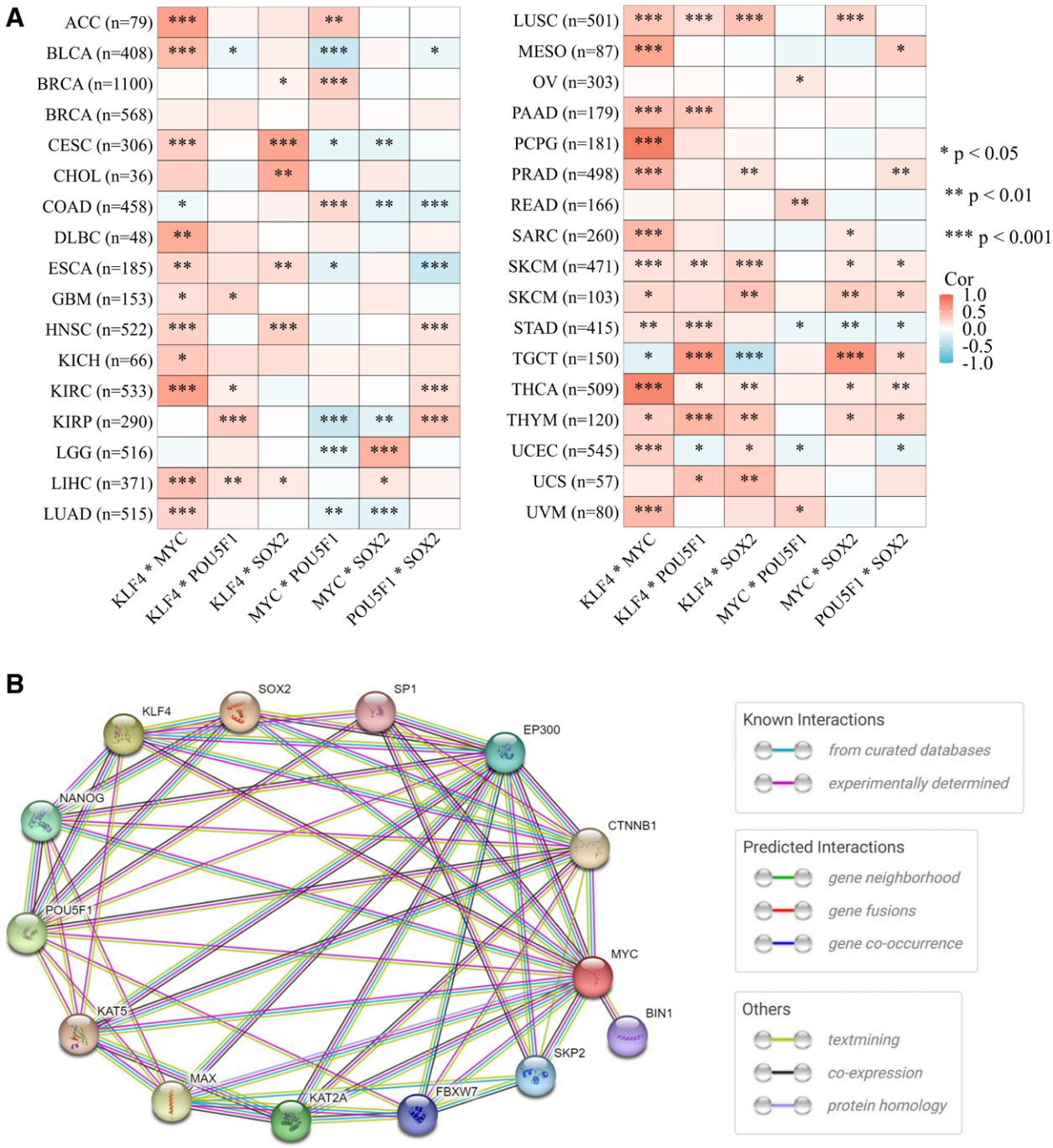

Figure 2. Expression correlation and protein-protein interaction of OSKM genes. (A) Heatmap of the relationship between mRNA transcription levels of OSKM genes. Spearman's correlation test were used to evaluate the degree of correlation. *$P < .05$, **$P < .01$, ***$P < .001$. (B) Protein-protein interaction networks of OSKM. OSKM = POU5F1 (OCT4), SOX2, KLF4 and MYC.

by POU5F1 in BLCA, COAD, and COADREAD, while induced in ACC, BRCA, KPAN, KIRP, NB, OV, THYM, UCEC, and UCS. SOX2 acts as negative regulator for immune checkpoint genes in LUSC, MESO, SARC, STES, and TGCT, while as positive regulator in PRAD and THYM. Collectively, these results suggest that OKM and SOX2 played different roles in immune infiltration and immune escape of cancers.

4. Discussion

The reprogramming of somatic cells to induced pluripotent stem cells (iPSCs) is a groundbreaking technique, which promises to deliver personalized treatments tailored to the unique needs of each patient. POU5F1 (OCT4), MYC, KLF4, and SOX2 have been identified as the 4 key transcription factors that play a central role in iPSC generation. Recent studies have shown that these factors function in a highly coordinated manner with each other, creating a synergistic effect that drives the reprogramming process. For example, KLF4 enhances the expression of OCT4 and SOX2,[39] while MYC promotes the establishment of pluripotency by promoting cellular proliferation.[40] Furthermore, SOX2 and OCT4 form a core circuit that regulates stem cell self-renewal and differentiation.[41]

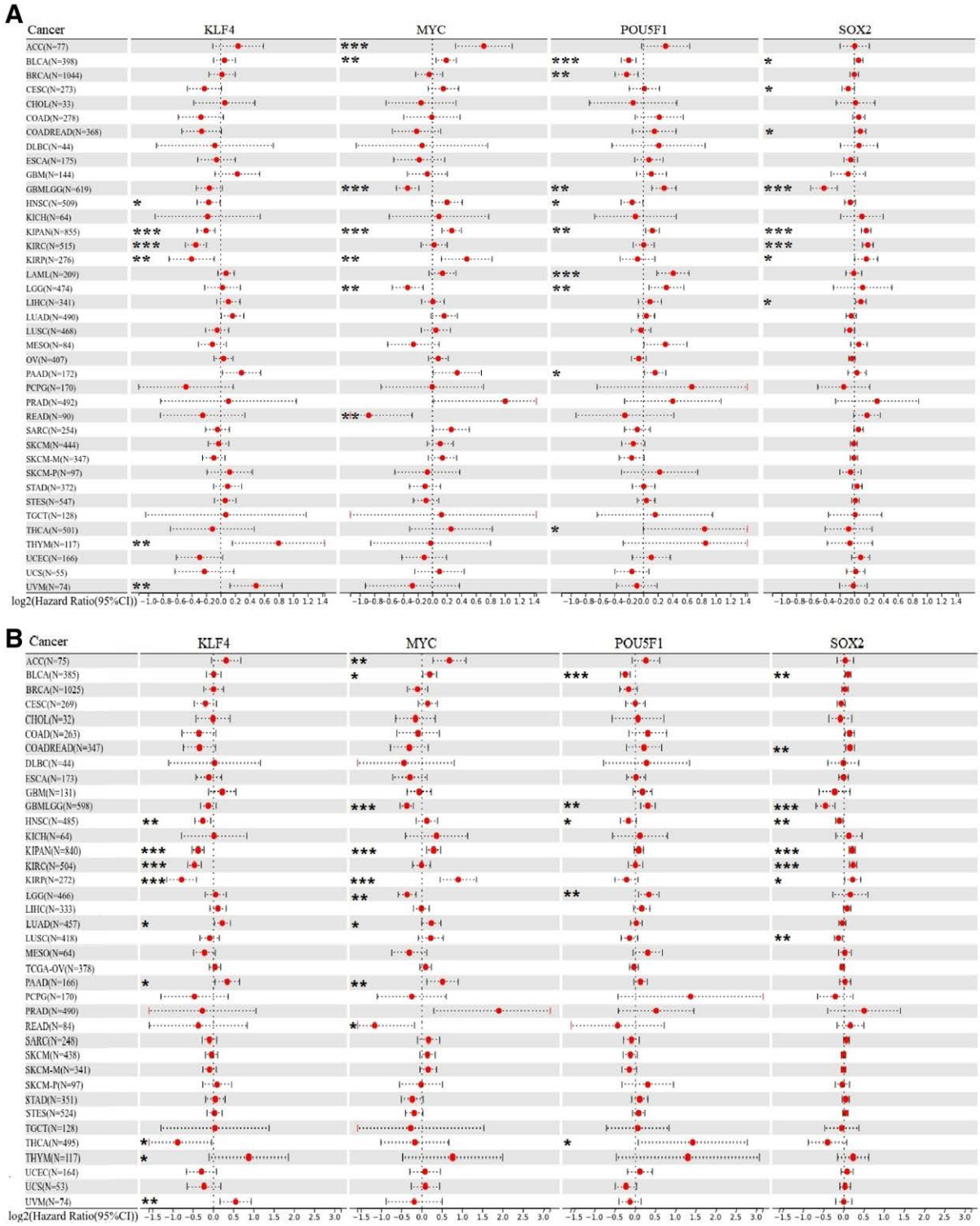

Figure 3. The survival analysis of OSKM expression in tumors. (A) Overall survival analysis. (B) Disease specific survival. Prognostic significance was obtained by statistical tests using Logrank test. *$P < .05$, **$P < .01$, ***$P < .001$. OSKM = POU5F1 (OCT4), SOX2, KLF4, and MYC.

The studies of stem cells contribute to the understanding of human growth and tumorigenesis, because the functional capabilities of normal stem cells and neoplastic cells are conceptually similar due to their ability to proliferate extensively. There are also high similarities between the dedifferentiation process of tumorigenesis and somatic cell reprogramming, which have given researchers new perspectives into how cancer develops and spreads as well as potential avenues for patient care. However, careful consideration of the similarity and differences between normal stem cells and tumorigenic cells is required for medical application for cancer treatment, tumor diagnosis and other clinical uses in the future. Uncovering the critical differences of molecular regulation between tumorigenesis and somatic cell reprogramming promises to elucidate fundamental aspects of dedifferentiation, which must

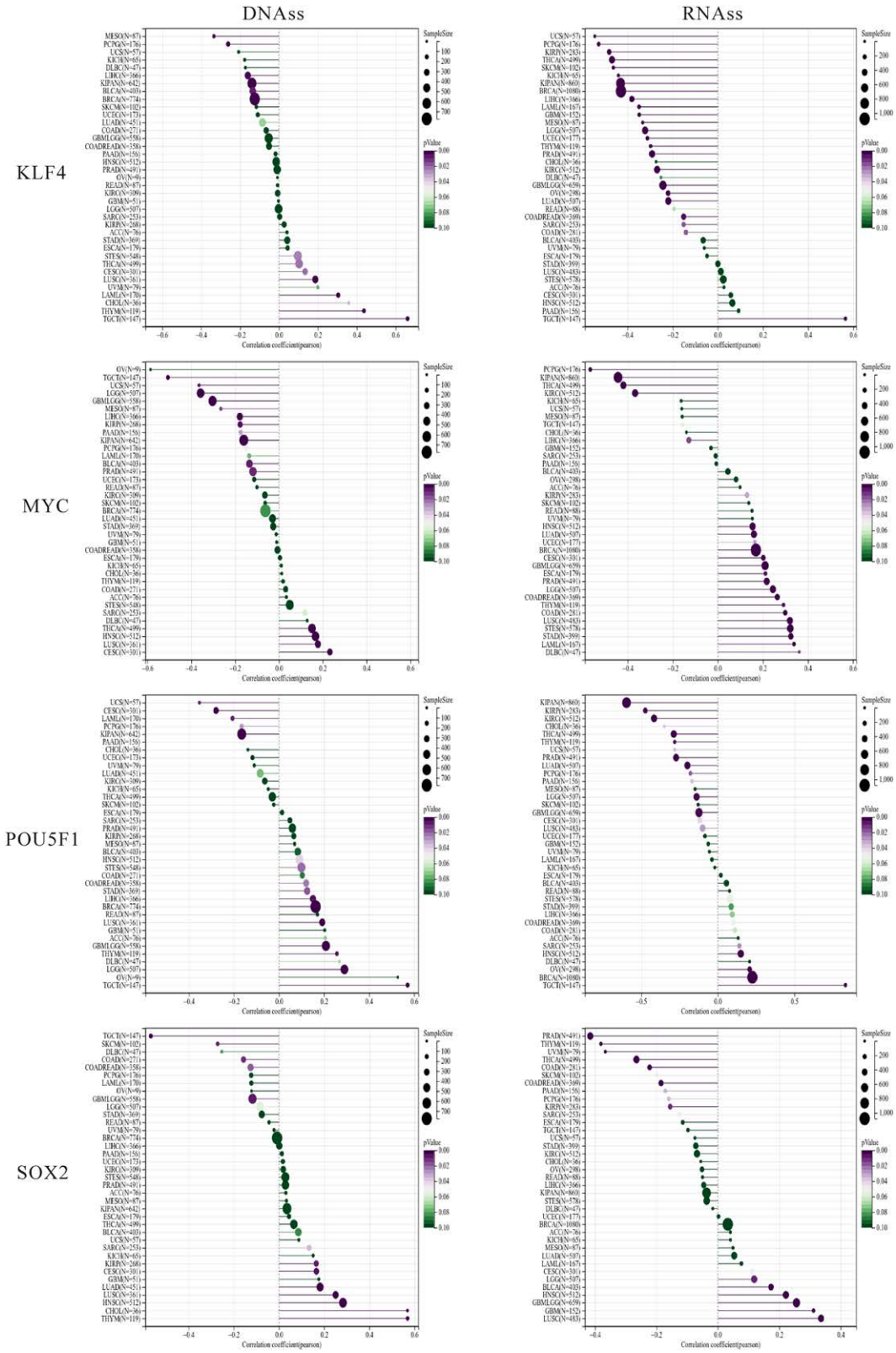

Figure 4. Correlation coefficients between OSKM expression and stemness scores DNAss and RNAss. Spearman's correlation test were used to study correlation between gene expression and stemness scores. The *P* values were represented by different colors as shown in the sidebar. OSKM = POU5F1 (OCT4), SOX2, KLF4 and MYC.

be essential for understanding the safety of cell replacement products derived from pluripotent stem cells.

To further understand the similarity and differences between stem cells and cancer cells, key reprogramming factors' function on pan-cancer were investigated using public resource databases in this study. The results showed that mRNA expression and protein levels of Yamanaka factors were detected in all tumor types with significant changes in most cancer tissues compared

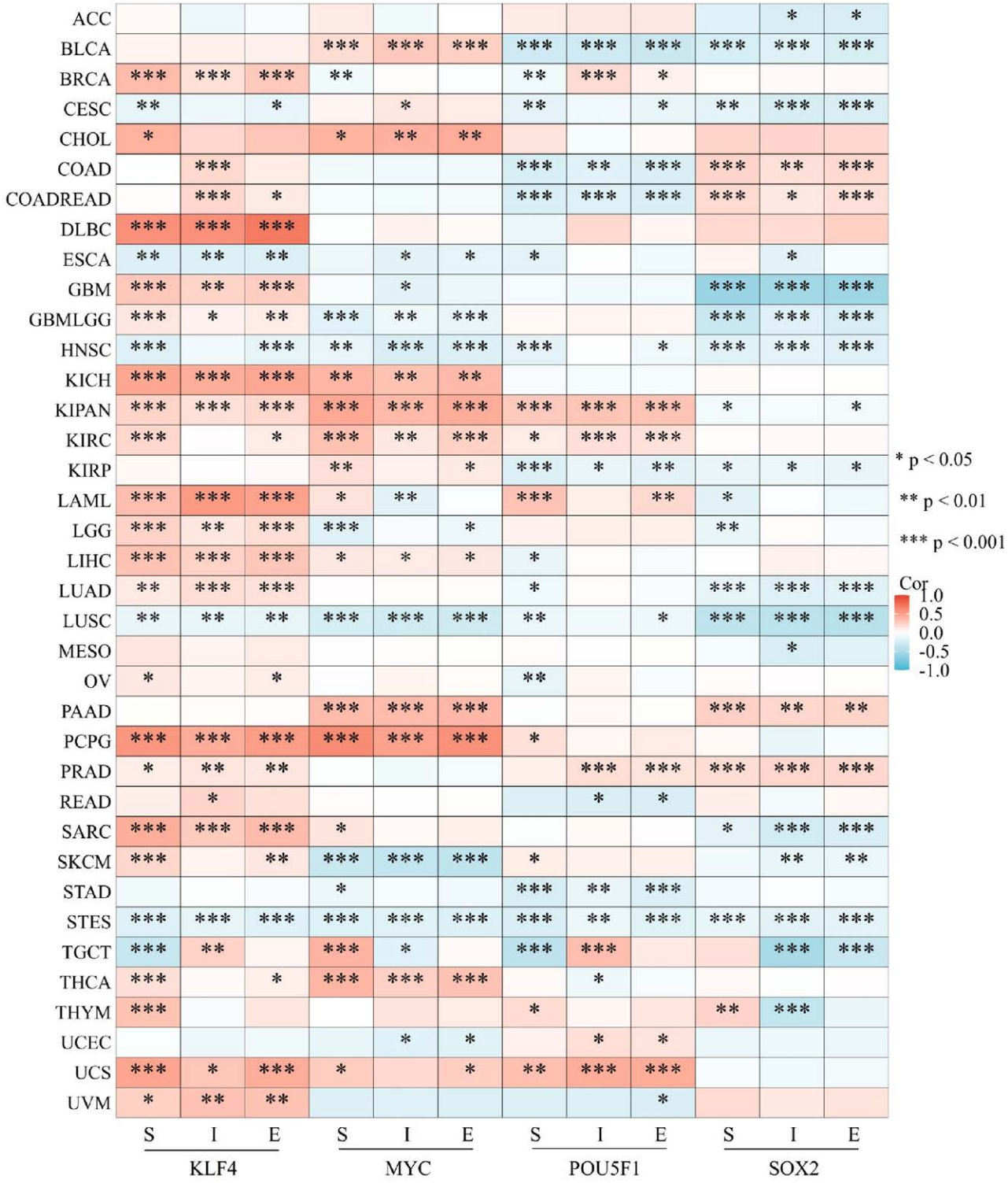

Figure 5. Correlation analysis of OSKM expression with the tumor immune microenvironment scores. S: stromal score, I: immune score, E: estimate score. Pearson's correlation coefficient was used to assess the association of OSKM genes with immune scores in individual tumors. *$P < .05$, **$P < .01$, ***$P < .001$. OSKM = POU5F1 (OCT4), SOX2, KLF4, and MYC.

with the respective normal tissues. The correlation of expression and protein-protein interaction were also verified among OSKM genes. Furthermore, the prognosis of patients, stemness of cancers cells, tumor microenvironment and immune checkpoint were affected by OSKM genes. All these data together indicates that the iPSC reprogramming factors play a role in tumorigenesis and spread.

However, when comparing the impact of each gene of OSKM on pan-cancer, their roles are diverse, in contrast to their synergistic regulation of iPSC reprogramming. For example, MYC showed positive correlation with tumor microenvironment and immune checkpoint genes, whereas SOX2 was negatively correlated with these processes. And OSKM genes all play a role in regulating tumor stemness, but it is difficult to generalize the

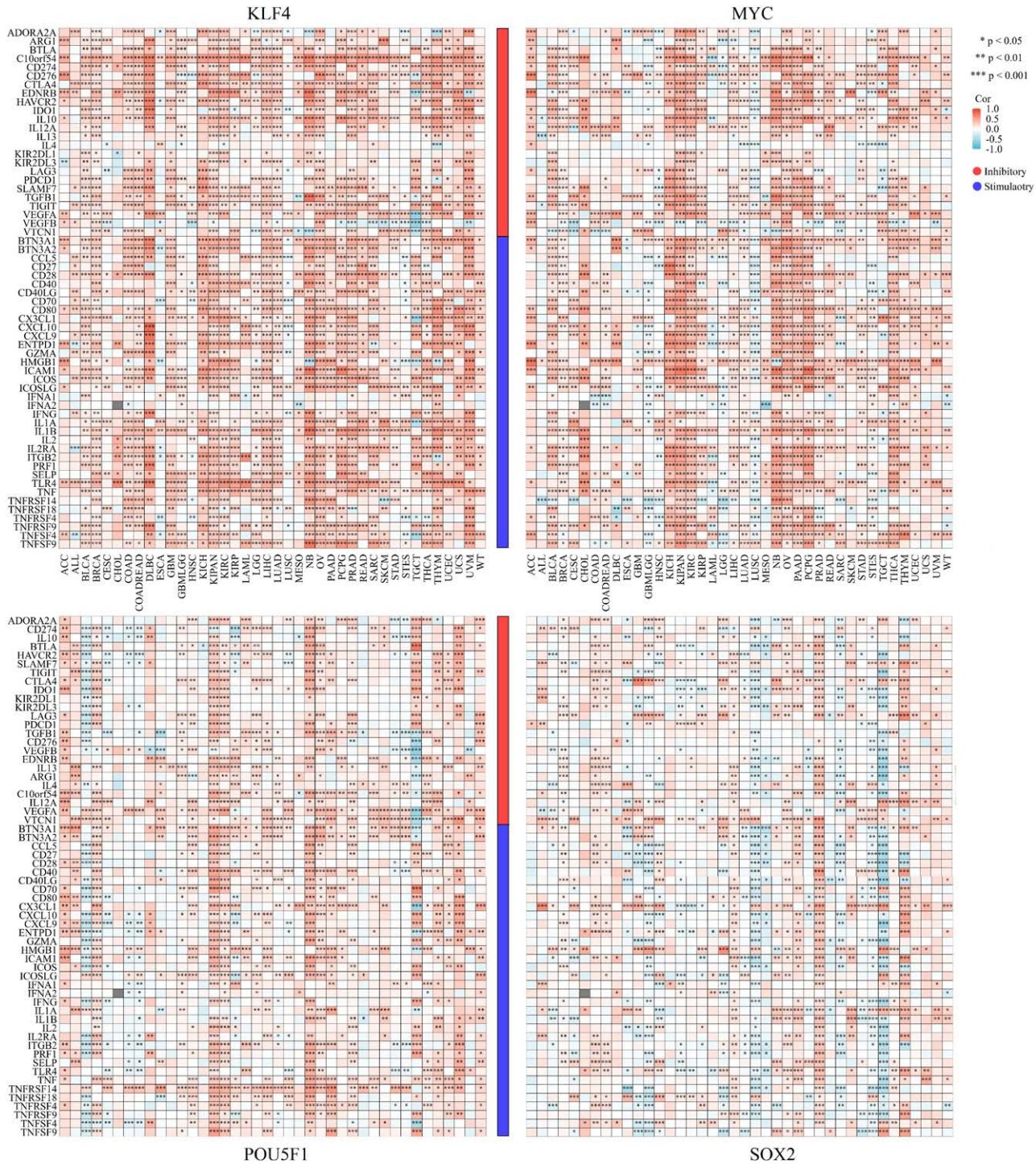

Figure 6. Correlation between transcription levels of OSKM and acknowledged immune checkpoints in multiple tumors. Red bar: 24 inhibitory genes, blue bar: 36 stimulatory genes. Spearman's correlation test were used to evaluate the correlation between OSKM and immune checkpoint genes. *$P < .05$, **$P < .01$, ***$P < .001$. OSKM = POU5F1 (OCT4), SOX2, KLF4, and MYC.

relevance between them from their DNAss and RNAss scores in different cancer types. By investigating the role of stemness factors in pan-cancer, this study promotes people's understanding of the similarities and differences between cancer cells and stem cells.

The differential expression and roles of OSKM genes in various cancers may arise for several reasons. Firstly, the regulation mechanisms of OSKM genes may vary across different cancers, including differences in gene regulatory networks and epigenetic modifications. Secondly, downstream factors that influence OSKM gene function may differ among cancer types. Moreover, the different effects of OSKM genes on the cellular microenvironment in different cancer types may in turn act on themselves. These together may be intrinsic to the similarities and differences in the occurrence of carcinogenesis and iPSC reprogramming processes.

Although studies investigating the differential expression and functions of OKSM genes in various cancers using publicly available databases have provided valuable insights, it is important to acknowledge the inherent limitations associated

with relying on these data. The clinical databases of these cancers consist of data derived from different experimental techniques and platforms, which may introduce inconsistencies and discrepancies in the data. Public databases are typically sourced from specific studies or cohorts, which could introduce potential sampling bias and may not represent the entire population of cancer cases. These can impact the accuracy and reliability of the analyses. Additionally, while public databases offer a vast sample size, they also present challenges in terms of systematic validation, making further experimental validation necessary to support the conclusions drawn in this study.

Acknowledgments

We thank Dr Congting Ye, as a senior bioinformatics practitioner of Xiamen University, for his careful review of the statistical analysis methods employed in this study.

Author contributions

Conceptualization: Liwei Hong.
Data curation: Liwei Hong, Sijie Hong, Xueqin Zhang.
Formal analysis: Liwei Hong, Sijie Hong, Xueqin Zhang.
Funding acquisition: Xueqin Zhang.
Investigation: Liwei Hong, Sijie Hong, Xueqin Zhang.
Methodology: Liwei Hong, Sijie Hong, Xueqin Zhang.
Project administration: Liwei Hong.
Resources: Liwei Hong, Sijie Hong.
Software: Liwei Hong, Sijie Hong, Xueqin Zhang.
Supervision: Liwei Hong.
Validation: Liwei Hong.
Visualization: Liwei Hong, Sijie Hong.
Writing – original draft: Liwei Hong.
Writing – review & editing: Liwei Hong, Sijie Hong, Xueqin Zhang.

References

[1] Deo SVS, Sharma J, Kumar S. GLOBOCAN 2020 report on global cancer burden: challenges and opportunities for surgical oncologists. Ann Surg Oncol. 2022;29:6497–500.
[2] Prasad S, Ramachandran S, Gupta N, et al. Cancer cells stemness: a doorstep to targeted therapy. Biochim Biophys Acta Mol Basis Dis. 2020;1866:165424.
[3] Tu SM. Cancer: a "stem-cell" disease? Cancer Cell Int. 2013;13:40.
[4] Wu B, Shi X, Jiang M, et al. Cross-talk between cancer stem cells and immune cells: potential therapeutic targets in the tumor immune microenvironment. Mol Cancer. 2023;22:38.
[5] Oshimori N, Guo Y, Taniguchi S. An emerging role for cellular cross-talk in the cancer stem cell niche. J Pathol. 2021;254:384–94.
[6] Chen L, Kasai T, Li Y, et al. A model of cancer stem cells derived from mouse induced pluripotent stem cells. PLoS One. 2012;7:e33544.
[7] Tapia N, Scholer HR. p53 connects tumorigenesis and reprogramming to pluripotency. J Exp Med. 2010;207:2045–8.
[8] Osman A, Afify SM, Hassan G, et al. Revisiting cancer stem cells as the origin of cancer-associated cells in the tumor microenvironment: a hypothetical view from the potential of iPSCs. Cancers. 2020;12:879.
[9] Raina K, Dey C, Thool M, et al. An insight into the role of UTF1 in development, stem cells, and cancer. Stem Cell Rev Rep. 2021;17:1280–93.
[10] Jha A. Tumor dedifferentiation: diagnostic and therapeutic implications. J Pathol Nepal. 2017;7:1202.
[11] Huang T, Song X, Xu D, et al. Stem cell programs in cancer initiation, progression, and therapy resistance. Theranostics. 2020;10:8721–43.
[12] Zhou Y, Li M, Zhou K, et al. Engineering induced pluripotent stem cells for cancer immunotherapy. Cancers (Basel). 2022;14:2266.
[13] Duinsbergen D, Salvatori D, Eriksson M, et al. Tumors originating from induced pluripotent stem cells and methods for their prevention. Ann N Y Acad Sci. 2009;1176:197–204.
[14] Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126:663–76.
[15] Takahashi K, Tanabe K, Ohnuki M, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007;131:861–72.
[16] Yu J, Vodyanik MA, Smuga-Otto K, et al. Induced pluripotent stem cell lines derived from human somatic cells. Science. 2007;318:1917–20.
[17] Wang L, Su Y, Huang C, et al. NANOG and LIN28 dramatically improve human cell reprogramming by modulating LIN41 and canonical WNT activities. Biol Open. 2019;8:bio047225.
[18] Getachew H, Chinchilla B, Fernandez-Godino R. Genome editing of induced pluripotent stem cells using CRISPR/Cas9 ribonucleoprotein complexes to model genetic ocular diseases. Methods Mol Biol. 2022;2549:321–34.
[19] Hong Y, Chan N, Begum AN. Deriving neural cells from pluripotent stem cells for nanotoxicity testing. Methods Mol Biol. 2019;1894:57–72.
[20] Hong Y, Dong X, Chang L, et al. Microglia-containing cerebral organoids derived from induced pluripotent stem cells for the study of neurological diseases. iScience. 2023;26:106267.
[21] Ben-David U, Benvenisty N. The tumorigenicity of human embryonic and induced pluripotent stem cells. Nat Rev Cancer. 2011;11:268–77.
[22] Metz EP, Wilder PJ, Popay TM, et al. Elevating SOX2 downregulates MYC through a SOX2: MYC signaling axis and induces a slowly cycling proliferative state in human tumor cells. Cancers (Basel). 2022;14:1946.
[23] Li K, Sun X, Zha R, et al. Counterintuitive production of tumor-suppressive secretomes from Oct4- and c-Myc-overexpressing tumor cells and MSCs. Theranostics. 2022;12:3084–103.
[24] Consortium ITP-CAoWG. Pan-cancer analysis of whole genomes. Nature. 2020;578:82–93.
[25] GTEx_Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45:580–5.
[26] Weinstein JN, Collisson EA, Mills GB, et al.; Cancer Genome Atlas Research Network. The cancer genome atlas pan-cancer analysis project. Nat Genet. 2013;45:1113–20.
[27] Shen W, Song Z, Zhong X, et al. Sangerbox: a comprehensive, interaction-friendly clinical bioinformatics analysis platform. Imeta. 2022;1:e36.
[28] Li T, Fu J, Zeng Z, et al. TIMER20 for analysis of tumor-infiltrating immune cells. Nucleic Acids Res. 2020;48:W509–14.
[29] Hu J, Yu A, Othmane B, et al. Siglec15 shapes a non-inflamed tumor microenvironment and predicts the molecular subtype in bladder cancer. Theranostics. 2021;11:3089–108.
[30] Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023;51:D638–46.
[31] Liu J, Lichtenberg T, Hoadley KA, et al.; Cancer Genome Atlas Research Network. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell. 2018;173:400–416.e11.
[32] Therneau T, Lumley T. R survival package. R Core Team. 2013.
[33] Malta TM, Sokolov A, Gentles AJ, et al.; Cancer Genome Atlas Research Network. Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell. 2018;173:338–354.e15.
[34] Yoshihara K, Shahmoradgoli M, Martinez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612.
[35] Primeaux M, Gowrikumar S, Dhawan P. Role of CD44 isoforms in epithelial-mesenchymal plasticity and metastasis. Clin Exp Metastasis. 2022;39:391–406.
[36] Mun J-Y, Leem S-H, Lee JH, et al. Dual relationship between stromal cells and immune cells in the tumor microenvironment. Front Immunol. 2022;13:864739.
[37] Li B, Chan HL, Chen P. Immune checkpoint inhibitors: basics and challenges. Curr Med Chem. 2019;26:3009–25.
[38] Sharma P, Allison JP. The future of immune checkpoint therapy. Science. 2015;348:56–61.
[39] Schmidt R, Plath K. The roles of the reprogramming factors Oct4, Sox2 and Klf4 in resetting the somatic cell epigenome during induced pluripotent stem cell generation. Genome Biol. 2012;13:251.
[40] Chappell J, Dalton S. Roles for MYC in the establishment and maintenance of pluripotency. Cold Spring Harb Perspect Med. 2013;3:a014381.
[41] Rizzino A, Wuebben EL. Sox2/Oct4: a delicately balanced partnership in pluripotent stem cells and embryogenesis. Biochim Biophys Acta. 2016;1859:780–91.