COMPUTATIONAL APPROACHES APPLIED TO CANCER GENETICS, IMMUNOGENOMICS, AND IMMUNO-ONCOLOGY

EDITED BY: Luis Zapata, Diego Chowell and Eszter Lakatos
PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-782-7
DOI 10.3389/978-2-83250-782-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

COMPUTATIONAL APPROACHES APPLIED TO CANCER GENETICS, IMMUNOGENOMICS, AND IMMUNO-ONCOLOGY

Topic Editors:
Luis Zapata, Institute of Cancer Research (ICR), United Kingdom
Diego Chowell, Icahn School of Medicine at Mount Sinai, United States
Eszter Lakatos, Queen Mary University of London, United Kingdom

Citation: Zapata, L., Chowell, D., Lakatos, E., eds. (2022). Computational Approaches Applied to Cancer Genetics, Immunogenomics, and Immuno-oncology. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83250-782-7

Table of Contents

frontiers
in Genetics

ORIGINAL RESEARCH
published: 27 September 2021
doi: 10.3389/fgene.2021.741226

Construction and Validation of an Immune Cell Signature Score to Evaluate Prognosis and Therapeutic Efficacy in Hepatocellular Carcinoma

Linfeng Xu[1,2†], Xingxing Jian[2,3†], Zhenhao Liu[2,3], Jingjing Zhao[2,4], Siwen Zhang[2,4], Yong Lin[1] and Lu Xie[2,3,4*]*

[1] School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai, China, [2] Shanghai Center for Bioinformation Technology, Shanghai Institute for Biomedical and Pharmaceutical Technologies, Shanghai, China, [3] Bioinformatics Center, National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China, [4] College of Food Science and Technology, Shanghai Ocean University, Shanghai, China

Edited by:
Eszter Lakatos,
Queen Mary University of London,
United Kingdom

Reviewed by:
Jinhui Liu,
Nanjing Medical University, China
Meiyi Li,
Fudan University, China

***Correspondence:**
Yong Lin
yong_lynn@163.com
Lu Xie
luxiex2017@outlook.com

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

***Received:** 14 July 2021*
***Accepted:** 30 August 2021*
***Published:** 27 September 2021*

Citation:
Xu L, Jian X, Liu Z, Zhao J, Zhang S, Lin Y and Xie L (2021) Construction and Validation of an Immune Cell Signature Score to Evaluate Prognosis and Therapeutic Efficacy in Hepatocellular Carcinoma. Front. Genet. 12:741226. doi: 10.3389/fgene.2021.741226

Background: Hepatocellular carcinoma (HCC) is the most common primary liver malignancy with high morbidity and mortality worldwide. Tumor immune microenvironment (TIME) plays a pivotal role in the outcome and treatment of HCC. However, the effect of immune cell signatures (ICSs) representing the characteristics of TIME on the prognosis and therapeutic benefit of HCC patients remains to be further studied.

Materials and methods: In total, the gene expression profiles of 1,447 HCC patients from several databases, i.e., The Cancer Genome Atlas (TCGA), International Cancer Genome Consortium, and Gene Expression Omnibus, were obtained and applied. Based on a comprehensive collection of marker genes, 182 ICSs were evaluated by single sample gene set enrichment analysis. Then, by performing univariate and multivariate Cox analysis and random forest modeling, four significant signatures were selected to fit an immune cell signature score (ICSscore).

Results: In this study, an ICSscore-based prognostic model was constructed to stratify HCC patients into high-risk and low-risk groups in the TCGA-LIHC cohort, which was successfully validated in two independent cohorts. Moreover, the ICSscore values were found to positively correlate with the current American Joint Committee on Cancer staging system, indicating that ICSscore could act as a comparable biomarker for HCC risk stratification. In addition, when setting the four ICSs and ICSscores as features, the classifiers can significantly distinguish treatment-responding and non-responding samples in HCC. Also, in melanoma and breast cancer, the unified ICSscore could verify samples with therapeutic benefits.

Conclusion: Overall, we simplified the tedious ICS to develop the ICSscore, which can be applied successfully for prognostic stratification and therapeutic evaluation in HCC. This study provides an insight into the therapeutic predictive efficacy of prognostic ICS, and a novel ICSscore was constructed to allow future expanded application.

Keywords: hepatocellular carcinoma, tumor immune microenvironment, immune cell signature, ICSscore, prognostic stratification, therapeutic evaluation

INTRODUCTION

Hepatocellular carcinoma (HCC) accounts for 75 to 85% of primary liver cancer, and is the sixth most common and fourth fatal malignancy globally, with 1- and 3-year survival rates of 20 and 5%, respectively, and a median survival of 8 months (Olsen et al., 2010; Bray et al., 2018). About two-thirds of patients with HCC are frequently diagnosed at advanced stages, being characterized by an aggressive clinical course (Llovet et al., 2016). Although multiple clinical strategies can be applied for HCC treatment, including surgical resection, liver transplantation, radiofrequency ablation, and chemotherapy, the efficacy is limited by high recurrence rate (Bruix and Sherman, 2011; Kuhlmann and Blum, 2013; Heimbach et al., 2018). Currently, the tumor–node–metastasis (TNM) system is still the gold standard for risk stratification of HCC patients (Liu et al., 2016). However, the recurrence and survival for HCC patients vary widely within each stage grouping (Park et al., 2020).

Emerging evidences showed that the tumor immune microenvironment (TIME) plays a key role in the tumor progression, recurrence, and metastasis (Nishida and Kudo, 2017; Kurebayashi et al., 2018). The differences in the composition and abundance of tumor-infiltrating lymphocytes (TILs), such as T cells, macrophages, dendritic cells, and associated fibroblasts, have been reported to influence the prognosis of HCC patients in different ways (Tang et al., 2019). For example, CD45RO+ memory T lymphocyte infiltration leads to a favorable clinical outcome in solid tumors, such as colorectal, gastric, and esophageal cancer, implicating that it is a valuable biomarker for prognostic prediction for human solid malignances (Gabrielson et al., 2016; Hu and Wang, 2017). Further understanding of TIME would provide more advanced prognostic and therapeutic biomarkers for HCC patients (Fu et al., 2019; Zhang et al., 2019). However, only a small number of TILs can be assessed, and the accuracy of applying TILs in predicting prognosis and treatment responding was still limited (Garnelo et al., 2017).

In this study, based on a comprehensive collection of marker genes attached to immune cell signatures (ICSs) from literatures, several HCC transcriptomic datasets were applied to quantify the ICS by single sample gene set enrichment analysis (ssGSEA). Subsequently, after performing univariate and multivariate Cox analysis and random forest modeling, four significant ICSs associated with prognosis were identified to construct an immune cell signature score (ICSscore). In several independent cohorts, the ICSscore was successfully validated to be associated with risk stratification of HCC patients, including tumor vs. normal samples, early- vs. advanced-staging samples, and treatment-responding vs. non-responding samples. Also, the unified ICSscore was validated successfully in other solid tumors, e.g., melanoma and breast cancer.

MATERIALS AND METHODS

Dataset Acquisition and Preprocessing

In this study, several gene expression datasets and the available clinical information of HCC were collected from several databases, including The Cancer Genome Atlas (TCGA), Gene Expression Omnibus, and International Cancer Genome Consortium (ICGC). Therein, in the microarray datasets (GSE14520, GSE96792, GSE109211, and GSE104580), we extracted the probe expression (log2 intensity) and probe annotation, respectively. When a gene was mapped by multiple probes, the expression of the gene was represented by the median of the multiple probes. In the RNA-seq datasets (TCGA-LIHC and ICGC LIRI-JP), we took the read counts to log2-transformation for normalization (Lian et al., 2018). In order to make the gene expression profiling comparable between different platforms, we then normalized with the scale method by using the limma package in R (Wang et al., 2021). Patients with follow-up time 0 or without follow-up were excluded from datasets. The available clinical characteristics of these samples are summarized in **Supplementary Table S1**.

The HCC datasets (GSE96792 and GSE109211) that received sorafenib treatment were obtained to assess the risk score in treatment-responding or non-responding patients. The HCC dataset (GSE104580) was used to predict therapeutic efficacy of transcatheter arterial chemoembolization (TACE). In addition, the breast cancer and malignant melanoma datasets (GSE20181 and GSE91061) were also downloaded to evaluate risk score and therapeutic effect (Riaz et al., 2017).

Immune Cell Signatures and Normalized Enrichment Score

In this study, a comprehensive collection of marker genes marked to 184 ICSs was referred from a literature (Wang S. et al., 2020), in which these ICSs and the corresponding marker genes were collected from diverse resources, including previous studies and database. To be specific, 25 signatures were collected from Bindea et al. (2013), 68 signatures were collected from the study of Wolf et al. (2014), 17 signatures were downloaded from the ImmPort database (Bhattacharya et al., 2014), 24 cell signatures were collected from the study of Miao et al. (2020), and 22, 10, and 10 signatures were collected from CIBERSORT (Newman et al., 2015), MCPcounter (Becht et al., 2016; R package, version 1.2.0) and imsig (Nirmal et al., 2018; Rpackage, version 1.1.3), respectively.

To quantify the 184 ICSs in each sample by a normalized enrichment analysis, the ssGSEA was implemented based on the gene expression matrix by using R package GSVA (version 1.36.3; Hänzelmann et al., 2013). Based on the expression of those given genes marked to each ICS, the ssGSEA produces an enrichment fraction, which represents the absolute enrichment degree in each sample. More detailed marked gene sets are listed in **Supplementary Table S2**. In this study, due to the lack of some marker genes in the transcriptomic profiles, only 182 ICSs were evaluated for subsequent analysis.

Construction of Immune Cell Signature Score

Since some ICSs with low variance may harm the convergence of hazard ratio (HR), and HR can be adjusted by magnifying the variance of some ICSs, we tried to increase the variance by scaling up ICS 10-fold or 100-fold for subsequent analysis. Based on the quantitative enrichment matrix of the ICS above, we first performed the univariable Cox proportional hazards regression analysis. 26 ICSs were selected with a significance of less than 0.01. Subsequently, a random forest algorithm (R package randomForestSRC, version 2.10.1) was used to narrow down feature selection (Breiman, 2001), in which we set the number of the nsplit as 100 in the variable hunting function (Ishwaran, 2015). The variable importance (VIMP) was used to measure the variation of the random forest model's prediction error rate. We selected the ICS with the VIMP of higher than 0.01. Here, only four ICSs were retained for subsequent analysis, i.e., CSR_Activated_15701700, CHANG_CORE_SERUM_RESPONSE_UP, Type_1_T_helper_cell, and TREM1_data.

On the basis of the four selected ICSs, multivariable Cox proportional hazards regression analysis was performed, and an ICSscore was constructed based on the quantitative enrichment matrix of the ICS and the corresponding regression coefficients as follows:

$$ICSscore = \sum_{i=1}^{4} \beta_i * ICS_i$$

Where ICS_i denotes the ith ICS and β_i represents the coefficient of ICS_i obtained from multivariate Cox regression analysis.

In this study, the ICSscore of each sample was calculated by the above formula. In each dataset, those patients were divided into high-risk or low-risk groups based on the median ICSscores in their respective datasets, in order to avoid the batch effect among the different datasets, especially RNA sequencing and microarray.

Comparison of Immune Cell Signature Score-Based Prognostic Model

Three published prognostic models (Wang Y. et al., 2020; Zhang et al., 2020; Liu P. et al., 2021) regarding HCC were taken to compare our model constructed in this study. The risk scores were calculated for each model, respectively. The differences in continuous score *p*-values and concordance index (C-index) from the univariate Cox analysis were compared, respectively.

Identification of Differentially Expressed Genes

According to the list of marker genes attached to the 184 ICS, we selected those genes attached to the four selected ICS. Here, a total of 435 unique genes were extracted. Subsequently, by using the R package limma (version 3.44.3), those genes with differential abundance were identified, which met the thresholds of absolute value of log2 fold change greater than 1 and the *p*-value less than 0.05 (Ritchie et al., 2015).

Machine Learning Classifier Algorithm

XGBoost is an optimization algorithm of gradient boosting decision tree, which is to gather many classification and regression tree models together to form a strong classifier (Jiang et al., 2021). To construct the classifier that could predict responders and non-responders in sorafenib treatment and TACE treatment, we applied the XGBoost algorithm (Python 3.8.3, package XGBoost version 1.3.0).

Statistical Analysis

In this study, all statistical analyses were implemented in R software (version 4.0.3). The Kaplan–Meier survival curve was visualized by using gsurvplot function implemented in the R package survminer (version 0.4.8) and log-rank test was used to compare the overall survival (OS), progression-free interval (PFI), disease-free interval (DFI), and disease-specific survival (DSS) between the different groups. Univariate Cox regression analysis was used to determine the significant features associated with OS, PFI, DFI, and DSS by calculating HR, 95% confidence interval (CI), and *p*-value between the different groups. Multivariate Cox regression analysis was used to assess the confounding risk score by several significant features. Receiver operating characteristic (ROC) analysis was used to evaluate the accuracy of prognostic model by using the R package survivalROC (version 1.0.3). The boxplot was visualized by using the R package ggpubr (version 0.4.0) and the nomogram and calibration plots were visualized by using the R package rms (version 6.0-1). Subgroup analysis was performed by the coxph function implemented in the R package survival (version 3.2-7) and the forest plot was generated by using the R package forestplot (version 1.10.1).

RESULTS

An Immune Cell Signature Score Was Constructed to Significantly Stratify Hepatocellular Carcinoma Patients

Here, 347 HCC samples with OS information in the TCGA-LIHC cohort were used as training dataset for prognostic model construction. First, based on the gene expression profiles and a list of genes marked to ICSs, only 182 ICS were able to be quantitatively evaluated. Subsequently, the evaluated ICSs were used to perform univariate Cox regression analysis, and 26 ICSs were selected with a *p*-value of less than 0.01 (**Supplementary Table S3**). To further narrow down features, we carried out dimension reduction analysis by using random forest algorithm, and four ICSs were identified with the VIMP of larger than 0.01, including CHANG_CORE_SERUM_RESPONSE_UP, CSR_Activated_15701700, TREM1_data, and Type_1_T_helper_cell (**Supplementary Figure S1A**). Eventually, the four selected ICSs were applied to construct a multivariate Cox prognostic model, in which an ICSscore was formulated based on the quantitative ICSs and their corresponding coefficients. The associations between the four ICSs and OS are illustrated in **Supplementary Figure S1B**, and the C-index of the prognostic model reached 0.70.

In order to examine whether ICSscore was an independent prognostic factor in each subgroup, ICSscore was applied to separately perform univariate Cox analysis in different subgroups of the TCGA-LIGC cohort, such as age, gender, American Joint

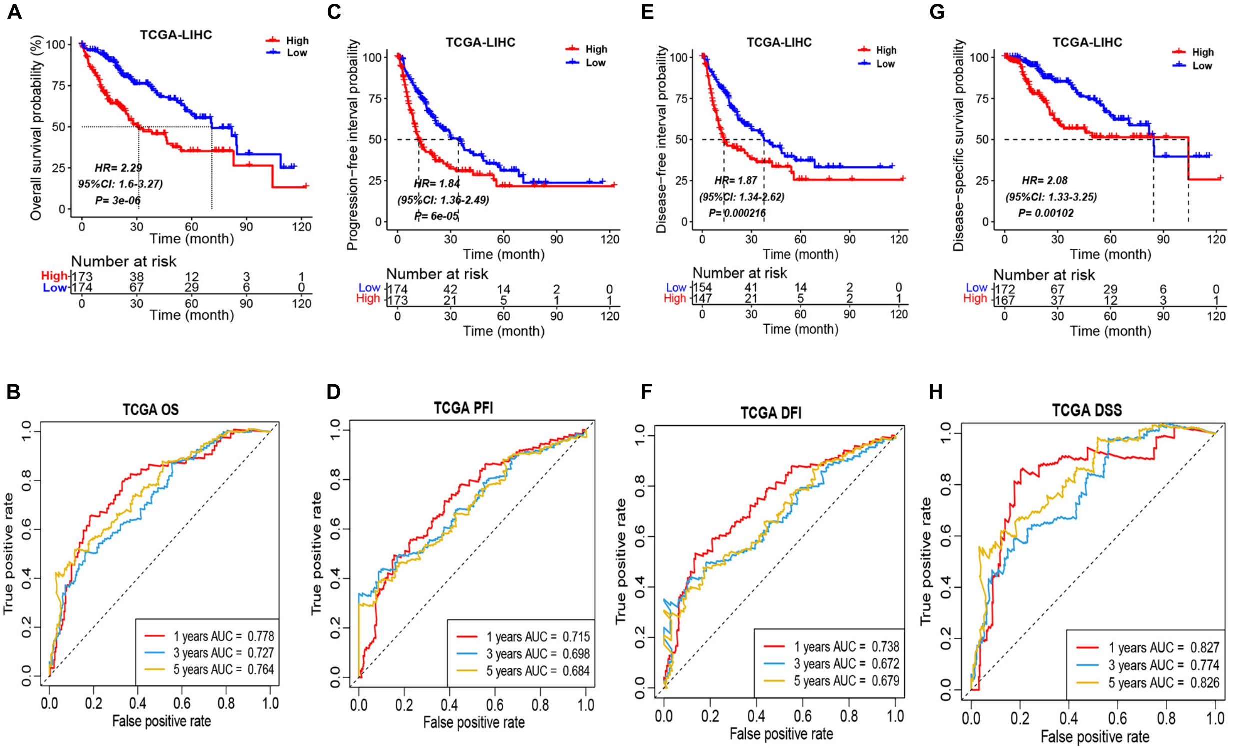


FIGURE 1 | Prognostic stratification of ICSscore in the TCGA-LIHC cohort. Patients were assigned to high-level and low-level groups by setting the median ICSscore as the cutoff. **(A)** The overall survival probability of high-level and low-level groups was evaluated (log-rank test, *p* = 3E-06). **(B)** The survival AUCs of 1-, 3-, and 5-year overall survival rate, respectively, were 0.778, 0.727, and 0.784. **(C)** The progression-free interval probability of high-level and low-level groups was evaluated (log-rank test, *p* = 6E-05). **(D)** The survival AUCs of 1-, 3-, and 5-year progression-free interval rate, respectively, were 0.715, 0.698, and 0.684. **(E)** The disease-free interval probability of high-level and low-level groups was evaluated (log-rank test, *p* = 0.000216). **(F)** The survival AUCs of 1-, 3-, and 5-year disease-free interval rate, respectively, were 0.738, 0.672, and 0.679. **(G)** The disease-specific survival probability of high-level and low-level groups was evaluated (log-rank test, *p* = 0.00102). **(H)** The survival AUCs of 1-, 3-, and 5-year disease-specific survival rate, respectively, were 0.827, 0.774, and 0.826.

Committee on Cancer (AJCC) stage, and vascular tumor cell type. As illustrated in **Supplementary Figure S2**, except for the AJCC stage IV, ICSscore could stratify HCC patients in the other subgroups significantly. However, in HCC patients of AJCC stage IV, the insignificance of ICSscore to stratify HCC patients may be due to the small sample size.

According to the median ICSscore in the TCGA-LIHC cohort, the patients can be divided into high-risk and low-risk groups. As shown in **Figure 1A**, the patients in the high-risk group showed significantly poorer OS than those in the low-risk group, indicating that high-level ICSscore is associated with worse outcomes. Furthermore, to assess the sensitivity and specificity of the ICSscore-based prognostic model, we performed ROC analysis. The area under curve (AUC) achieved 0.778, 0.727, and 0.764, respectively, at the 1-, 3-, and 5-year OS rate (**Figure 1B**), suggesting that the ICSscore-based prognostic model has a good prediction performance.

Moreover, the differences in PFI, DFI, and DSS between the high-risk and low-risk groups in the TCGA-LIHC cohort were also compared, respectively. Consistently, the patients in the high-risk group all showed obviously poorer PFI (**Figure 1C**), DFI (**Figure 1E**), and DSS (**Figure 1G**). Meanwhile, through performing the ROC analysis on PFI, DFI, and DSS, the comparable AUCs are shown in **Figures 1D,F,H**. These implied that the ICSscore constructed by the four significant ICS can significantly stratify HCC patients.

To provide a clinically applicable risk assessment model for predicting the prognosis of HCC patients, a nomogram that integrated ICSscore and AJCC staging was constructed in the TCGA-LIHC cohort (**Supplementary Figure S3A**). According to the nomogram illustrated in this study, a combination of ICSscore and AJCC stage of a HCC patient can be calculated to predict the 1-, 3-, and 5-year OS for an individual. In addition, as illustrated in **Supplementary Figures S3B–D**, the calibration curves at the 1-, 3- and 5-year OS for an individual all fit well to the ideal curves. Noteworthy, we found that the ICSscore contributed to the most risk points when compared with the AJCC staging, suggesting that ICSscore would make a greater predictive contribution.

The Validation of the Immune Cell Signature Score-Based Prognostic Model

In order to validate the robustness of the ICSscore-based prognostic model trained in the TCGA-LIHC cohort, two

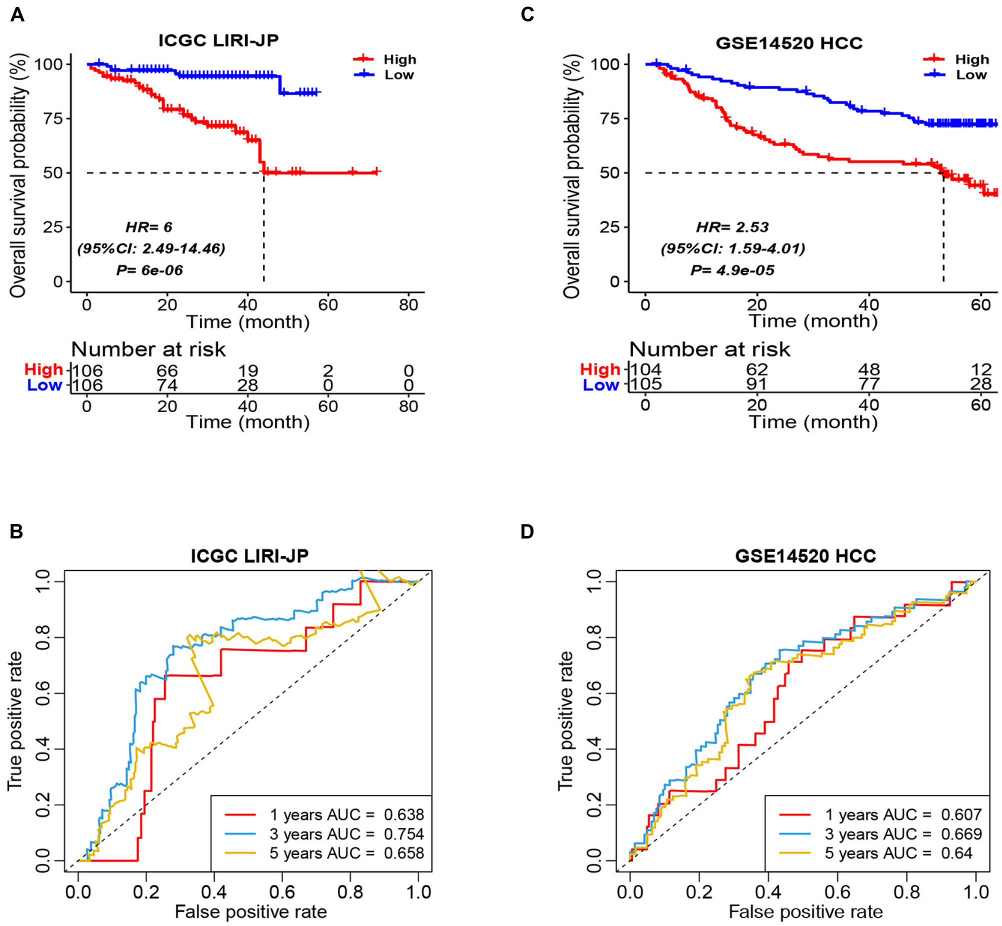


FIGURE 2 | Prognostic stratification of ICSscore in the ICGC LIRI-JP and GSE14520 HCC cohort. Patients were assigned to high-level and low-level groups by setting the respective median ICSscore as the cutoff. **(A)** The overall survival probability of high-level and low-level groups was evaluated in the ICGC LIRI-JP cohort (log-rank test, *p* = 6E-06). **(B)** The survival AUCs of 1-, 3-, and 5-year overall survival rate, respectively, were 0.638, 0.754, and 0.658. **(C)** The overall survival probability of high-level and low-level groups was evaluated in the GSE14520 HCC cohort (log-rank test, *p* = 4.9E-05). **(D)** The survival AUCs of 1-, 3-, and 5-year overall survival rate, respectively, were 0.607, 0.669, and 0.64.

independent datasets (i.e., ICGC LIRI-JP and GSE14520 HCC) were applied, respectively. Similarly, in the two validation cohorts, according to their individual median ICSscore, we divided patients into two groups, i.e., high-ICSscore and low-ICSscore groups. Consistent with the findings above, the high-level ICSscore group showed significantly poorer prognostic outcomes (**Figures 2A,C**). Meanwhile, the ROCs were also analyzed in the two validation cohorts. The AUC of the prognostic model was 0.638, 0.754, and 0.658, respectively, at 1-, 3-, and 5-year survival rates in the IGCG LIRI-JP cohort (**Figure 2B**), and the AUC was 0.607, 0.669, and 0.640, respectively, at 1-, 3-, and 5-year survival rates in the GSE14520 HCC cohort (**Figure 2D**). These results demonstrated that the ICSscore can be used to stratify HCC patients and predict prognosis.

In addition, in light of the ICSscore in different subgroups of the ICGC LIRI-JP cohort, we separately carried out univariate Cox analysis, such as age, gender, TNM stage, virus, and vein invasion. Except for those subgroups with small sample sizes, ICSscore did stratify significantly HCC patients (**Supplementary Figure S4**), suggesting that ICSscore was a robust biomarker to stratify patients in the different subgroups.

Furthermore, in the ICGC LIRI-JP cohort, a nomogram that integrated ICSscore and TNM staging was constructed as well (**Supplementary Figure S5A**). Compared with TNM staging, we also observed that the ICSscore contributed to the most risk points, demonstrating that the ICSscore can make a greater predictive contribution. Meanwhile, the calibration curves at the 1-, 3- and 5-year OS were all found to be close to the ideal curves (**Supplementary Figures S5B–D**).

The Comparison of Risk Stratification and Predictive Ability of Immune Cell Signature Score as a Feature

To compare the risk stratification and predictive ability of ICSscore, we calculated the continuous prognostic risk scores and concordance index (C-index) by performing univariate Cox analysis. Compared with age, gender, AJCC stage, and invasion (**Tables 1**, **2**), the C-index of the ICSscore was higher and the *p*-value of the ICSscore was lower, indicating ICSscore to be a good predictor.

In addition, three published prognostic models (Wang Y. et al., 2020; Zhang et al., 2020; Liu P. et al., 2021) regarding HCC were used to compare our ICSscore-based prognostic model constructed in this study. The continuous prognostic risk scores were calculated for each model by performing univariate Cox analysis, respectively, in TCGA-LIHC and ICGC-JP cohorts. As shown in **Tables 1**, **2**, these differences in *p*-values and C-index were compared, suggesting that our ICSscore-based prognostic model has a preferrable predictive ability.

Differential Marker Genes in the Four Immune Cell Signatures Formulating Immune Cell Signature Score

To explore the underlying reason of ICSscore in risk assessment and prognostic prediction, 435 marker genes attached to the four significant ICS formulating ICSscore were investigated. In the TCGA-LIHC cohort, the gene expression matrix from 347 tumor samples and 49 normal samples was used for subsequent differential analysis. First, between the tumor and normal samples, a total of 97 differentially expressed genes (DEGs) were identified, including 36 up-regulated genes and 61 down-regulated genes (**Figure 3A**). Similarly, between the high-risk and low-risk samples as distinguished above, we obtained 21 DEGs, including 11 up-regulated genes and 10 down-regulated genes (**Figure 3B**), which were speculated to make more contribution to differential ICSscore evaluation. Thus, we performed GO enrichment and several significant biological processes were obtained (**Supplementary Figure S6**), such as activated T-cell proliferation, positive regulation of wound healing, and regulation of activated T-cell proliferation. In addition, of these 21 genes, 15 were found to be the same as those between tumor and normal samples (**Figure 3C**). Notably, sequentially comparing the normal samples, the low-risk samples, and the high-risk samples, we found that the abundance of genes *FLNC*, *HAVCR1*, *PLK4*, *WDHD1*, *CENPW*, *MYBL2*, and *SKA1* increased, while genes *IGF2*, *SELP*, *GREM2*, *HSD11B1*, *CFHR3*, *GPLD1*, *F12*, and *PLG* decreased. Furthermore, univariate analysis of these genes showed that the upregulated genes were detrimental to HCC prognosis, while the down-regulated genes were beneficial (**Figure 3C**). Indeed, most of these genes have been reported as prognostic biomarkers or suggested as novel therapeutic targets for HCC. For example, the overexpression of genes *CENPW*, *MYBL2*, and *SKA1* is associated with poor prognosis in HCC, while the loss of gene *HSD11B1* indicates poor prognosis in HCC (Frau et al., 2011; Chen et al., 2018; Zhou et al., 2020). Moreover, we focused on the correlations between the ICSscore value and expression levels of 15 genes. As illustrated in **Figure 3D**, the up-regulated genes were positively correlated with the ICSscore, while the down-regulated genes were negatively correlated. Moreover, in the ICGC-JP cohort, as shown in **Figure 3E**, the abundance alteration of the above 15 genes and their association with prognosis were observed to be consistent.

TABLE 1 | Comparison of the *p*-value and C-index derived from the univariate Cox model in the TCGA-LIHC cohort.

Signatures	*p*-value	C-index
ICSscore	1.82E-13	0.700
Baohui_Zhang_2020	8.89E-10	0.694
Yu_Wang_2020	3.08E-13	0.690
Peng_Liu_2021	9.56E-09	0.640
Age	0.1881	0.508
Gender	0.2614	0.507
AJCC_STAGE	1.52E-05	0.609
Vascular_tumor_cell_type	0.123	0.533

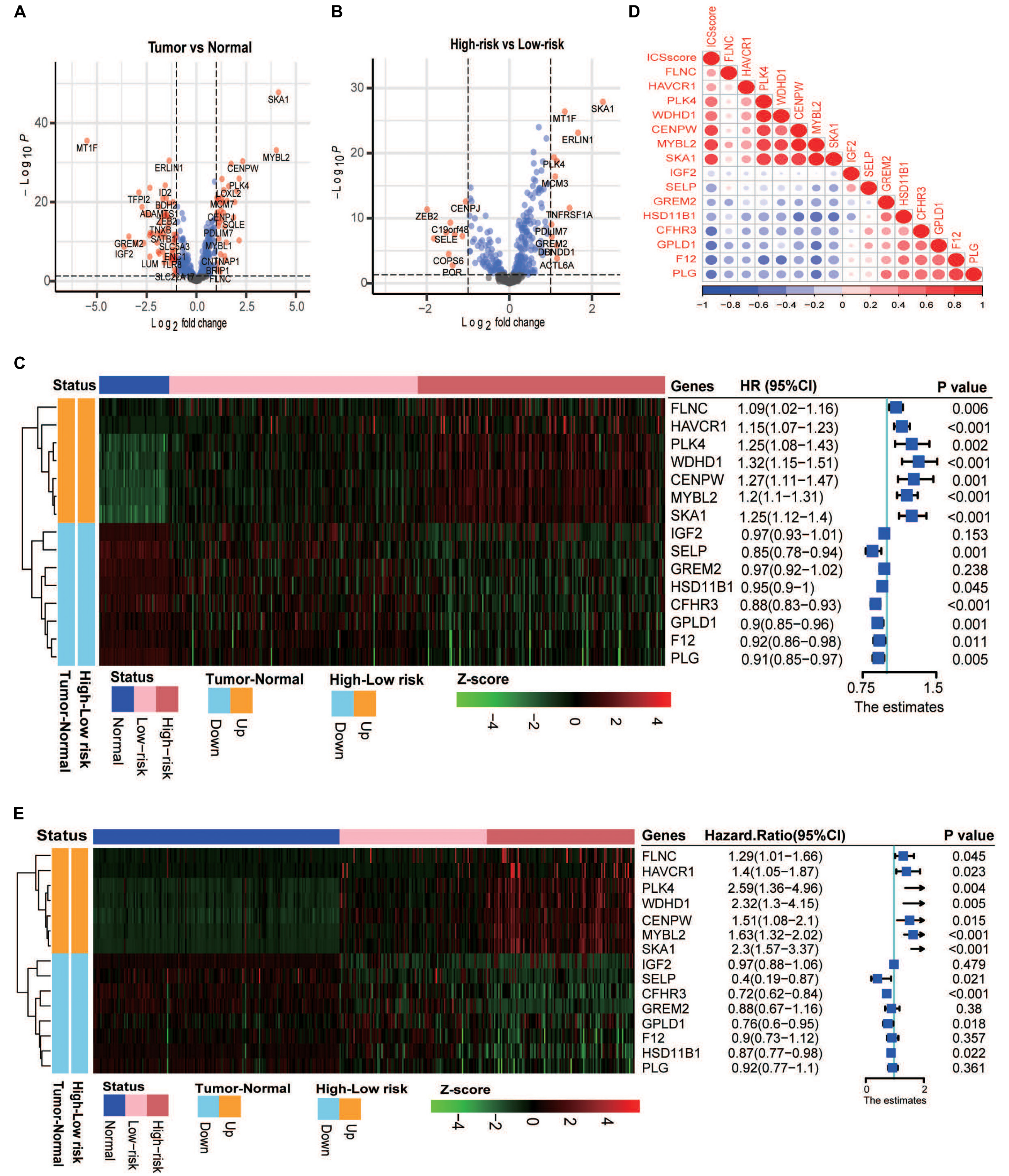
TABLE 2 | Comparison of the *p*-value and C-index derived from the univariate Cox model in the ICGC-JP cohort.

Signatures	*p*-value	C-index
ICSscore	9.44E-05	0.711
Baohui_Zhang_2020	7.75E-5	0.707
Yu_Wang_2020	0.0001582	0.680
Peng_Liu_2021	0.00285	0.671
Age	0.3973	0.536
Gender	0.07557	0.566
TNM_STAGE	0.0001536	0.704
VEIN_INVASION	0.004346	0.615

Evaluation and Prediction of Disease Malignancy and Molecular Target Therapy Benefit in Hepatocellular Carcinoma by Immune Cell Signature Score

In order to verify whether the ICSscore evaluation was consistent with other risk stratification methods, several HCC cohorts were compared. First, as illustrated in **Figure 4A**, in the three HCC cohorts, tumor samples all exhibited strikingly higher ICSscore values when compared with the paired normal samples. In the GSE25097 HCC cohort, we also found that the tumor samples showed the highest ICSscore values, while the normal samples showed relatively low ICSscore values, although there was no significant difference between the normal samples and cirrhotic samples (**Figure 4B**). These results indicate a significant increase in ICSscore when hepatocytes develop into tumors.

Recently, the eighth edition staging system of the AJCC was released for HCC stratification (Park et al., 2020). In the TCGA-LIHC cohort, after excluding three stage IV samples, the stage III samples showed the highest ICSscore, followed by stage II samples, and the stage I samples exhibited the lowest ICSscore, indicating that the ICSscore was positively correlated with the current risk stratification system (**Figure 4C**). As expected, we found that most advanced-staging patients (stage III and stage IV) were assigned into the high-risk group, while more early-staging patients (stage I and II) were designated into the low-risk group (**Figure 4D**), implying that the ICSscore could act as a comparable marker for HCC risk stratification. At the same



Genes	HR (95%CI)	P value
FLNC	1.09(1.02−1.16)	0.006
HAVCR1	1.15(1.07−1.23)	<0.001
PLK4	1.25(1.08−1.43)	0.002
WDHD1	1.32(1.15−1.51)	<0.001
CENPW	1.27(1.11−1.47)	0.001
MYBL2	1.2(1.1−1.31)	<0.001
SKA1	1.25(1.12−1.4)	<0.001
IGF2	0.97(0.93−1.01)	0.153
SELP	0.85(0.78−0.94)	0.001
GREM2	0.97(0.92−1.02)	0.238
HSD11B1	0.95(0.9−1)	0.045
CFHR3	0.88(0.83−0.93)	<0.001
GPLD1	0.9(0.85−0.96)	0.001
F12	0.92(0.86−0.98)	0.011
PLG	0.91(0.85−0.97)	0.005

Genes	Hazard.Ratio(95%CI)	P value
FLNC	1.29(1.01−1.66)	0.045
HAVCR1	1.4(1.05−1.87)	0.023
PLK4	2.59(1.36−4.96)	0.004
WDHD1	2.32(1.3−4.15)	0.005
CENPW	1.51(1.08−2.1)	0.015
MYBL2	1.63(1.32−2.02)	<0.001
SKA1	2.3(1.57−3.37)	<0.001
IGF2	0.97(0.88−1.06)	0.479
SELP	0.4(0.19−0.87)	0.021
CFHR3	0.72(0.62−0.84)	<0.001
GREM2	0.88(0.67−1.16)	0.38
GPLD1	0.76(0.6−0.95)	0.018
F12	0.9(0.73−1.12)	0.357
HSD11B1	0.87(0.77−0.98)	0.022
PLG	0.92(0.77−1.1)	0.361

FIGURE 3 | Analysis of differential gene expression in the TCGA-LIHC and ICGC-JP cohort. **(A)** Volcano plot presents the differentially expressed genes (DEGs) between the tumor and normal samples. **(B)** Volcano plot presents the DEGs between the high-risk and low-risk samples. **(C)** Left: heatmap shows the scaled abundance of 15 DEGs among normal, low-risk, and high-risk samples. Right: forest plot denotes the association between the DEGs and overall survival. The HR, 95% CI, and *p*-value were determined by univariate Cox regression analysis in the TCGA-LIHC cohort. **(D)** Correlations between expression levels of 15 genes and the ICSscore values. **(E)** Left: heatmap shows the scaled abundance of 15 DEGs among normal, low-risk, and high-risk samples. Right: forest plot denotes the association between the DEGs and overall survival. The HR, 95% CI, and *p*-value were determined by univariate Cox regression analysis in the ICGC-JP cohort.

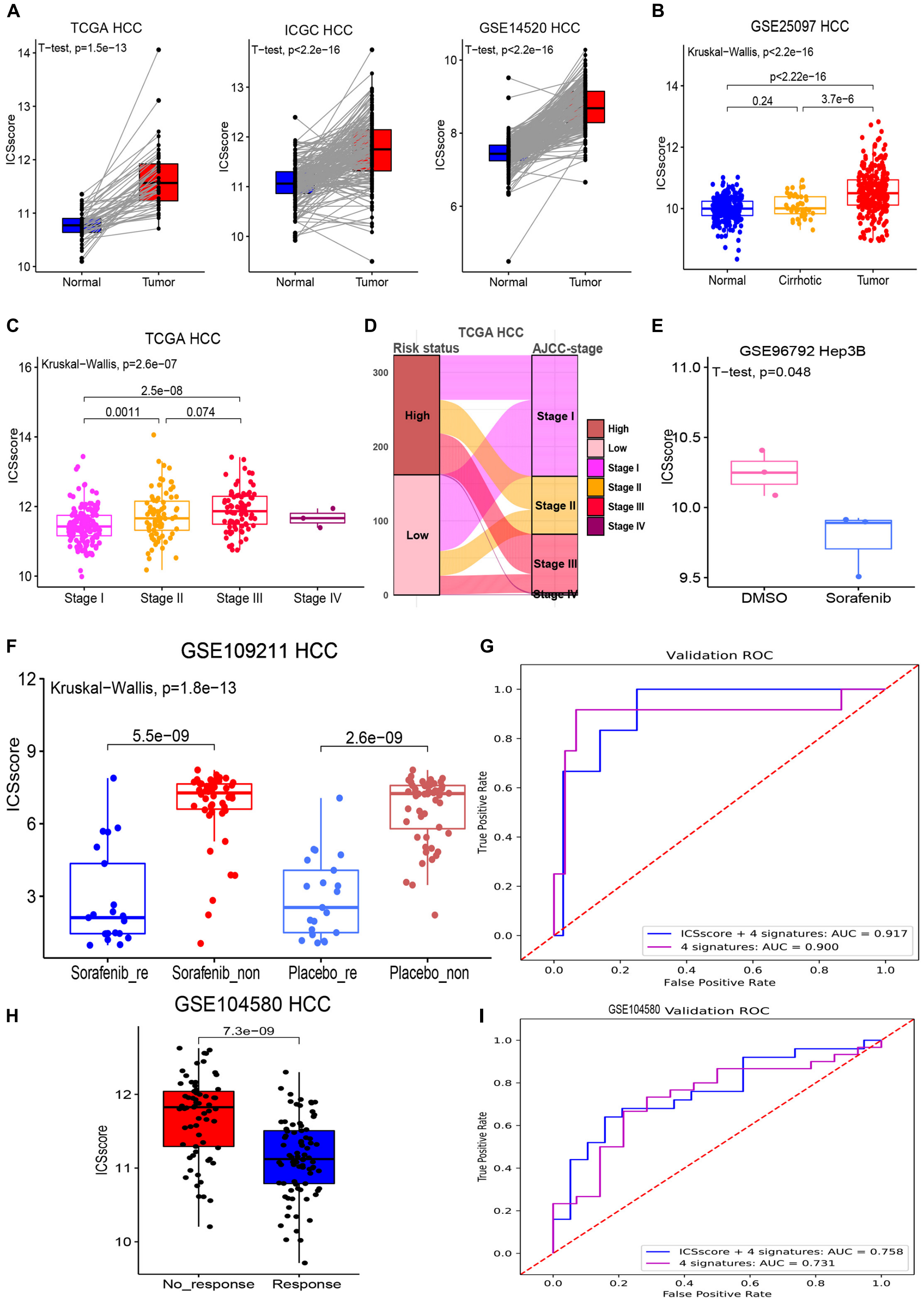


FIGURE 4 | Evaluation and prediction of disease malignancy and molecular target therapy benefit in HCC by ICSscore. **(A)** Pairwise comparison of the ICSscore between normal and tumor samples in three cohorts, i.e., TCGA-LIHC (*t*-test p = 1.5E-13), ICGC LIRI-JP (*t*-test $p < 2.2E-13$), and GSE14520 HCC (*t*-test $p < 2.2E-13$). **(B)** Boxplot illustrates the differences of the ICSscore values among normal, cirrhotic, and tumor samples in the GSE25097 HCC cohort. **(C)** Boxplot illustrates the differences of the ICSscore values among different AJCC staging of the TCGA-LIHC cohort. **(D)** Sankey plot shows the mapping between high or low ICSscore and AJCC staging of the TCGA-LIHC cohort. **(E)** Boxplot shows the ICSscore values of Hep3B cell line treated with sorafenib or DMSO in the GSE96792 cohort. **(F)** Boxplot illustrates the ICSscore values of responded or non-responded HCC patients treated with sorafenib or placebo in the GSE109211 cohort. **(G)** ROC curve of the XGBoost algorithm for predicting the responding and non-responding patients in the GSE109211 cohort. **(H)** Boxplot illustrates the ICSscore values of responding or non-responding HCC patients treated with chemotherapy in the GSE104580 cohort. **(I)** ROC curve of the XGBoost algorithm for predicting the responding and non-responding patients in the GSE104580 cohort.

time, we also observed that some early-staging patients were assigned to the high-risk group, while late-staging patients were assigned to the low-risk group, suggesting that ICSscore may be used as a supplement to compromise AJCC-staging risk stratification errors. Indeed, studies have reported significant differences in recurrence and survival for HCC patients within each AJCC stage grouping.

Furthermore, we explored whether ICSscore could be used as a marker to evaluate therapeutic efficacy. Sorafenib is the only Food and Drug Administration-approved first-line targeted agent for the treatment of advanced HCC, but its impact on patient survival is limited depending on the pathogenetic conditions (Bruix et al., 2017). Here, the gene expression profiles (GSE96792) from the Hep3B cell line treated with sorafenib or DMSO was obtained to evaluate the ICSscore, respectively. As a result, we did observe lower ICSscore in those Hep3B treated with sorafenib compared with those Hep3B treated with DMSO (**Figure 4E**), suggesting that the ICSscore may be used to reflect therapeutic efficacy. Subsequently, we further tested the ICSscore in a clinical trial on sorafenib (GSE1090211), and those patients who responded to sorafenib showed much lower ICSscore than those who had no response to sorafenib (Figure metricconverterProductID4F4F). Interestingly, in those patients treated with placebo, we also observed that responding patients exhibited much lower ICSscore than non-responding patients (Figure metricconverterProductID4F4F). These results implied that the ICSscore could be used to predict the therapeutic benefit. Therefore, when setting the four ICSs and ICSscores as features, two classification models were separately constructed to predict responding and non-responding samples, in which 70% of the data in the GSE109211 cohort was taken as the training set, and 30% as the validation set. The AUCs for predicting treatment responding were achieved at 0.917 and 0.900, respectively, (**Figure 4G**). Similarly, those HCC patients who received chemotherapy in GSE104580 cohorts were examined as well. Consistently, compared with those HCC patients who had no response to chemotherapy, much lower ICSscores were observed in those patients who responded to chemotherapy (**Figure 4H**). Also, when setting the four ICSs and ICSscores as features to build classification models, the AUCs for predicting treatment-responding patients were achieved at 0.758 and 0.733, respectively, (**Figure 4I**). These findings implied that the ICSscore may be used as an indicator for prediction of treatment responding in HCC.

Evaluation and Prediction of Chemotherapy and Immunotherapy Benefit in Other Tumors by Immune Cell Signature Score

As most patients with a high-level ICSscore displayed poorer prognosis and low therapeutic benefit than those with a low-level ICSscore in HCC, we explored whether the ICSscore could predict therapeutic benefit in other tumors. For this investigation, a cohort of breast cancer patients with chemotherapy information (GSE20181) were first applied to calculate the unified ICSscore value for each patient on the basis of transcriptomic profiles and marker genes. When comparing the pairwise ICSscore before and after treatment with adjuvant chemotherapy, we found that the patients' ICSscore significantly decreased after 14-day adjuvant chemotherapy (**Figure 5A**), and it decreased further after 90-day adjuvant chemotherapy (**Figure 5B**). That is to say, further adjuvant chemotherapy led to a gradual decrease of ICSscore, suggesting a gradual therapeutic benefit. This suggests that the ICSscore could be used to monitor therapeutic efficacy in breast cancer.

More recently, the strategy for immune checkpoints, PD-1 and PD-L1, has become an immune therapy with amazing survival benefit (Liu C. et al., 2021). Unfortunately, the effectiveness of immune checkpoint therapy is limited because only a small number of patients respond to the therapy. Here, a cohort of melanoma patients who received anti-PD1 and anti-CTLA4 therapy (GSE91016) were also applied to evaluate the ICSscore application. By setting the mean ICSscore value as the cutoff, these patients were classified into high-ICSscore and low-ICSscore groups. Similarly, the high-ICSscore group exhibited significantly poorer OS (**Figure 5C**). In addition, by pairwise comparing the ICSscore between patients before and after immunotherapy, we observed that the patients' ICSscore was significantly decreased after receiving immunotherapy (**Figure 5D**). The result implied that lower ICSscore values can be used to distinguish those patients who benefit from immunotherapy. Indeed, as shown in **Figure 5E**, the patients with CR/PR presented lower ICSscore than those with PD/SD. Subsequently, setting the four ICSs and ICSscores as features, we constructed two classification models to predict whether the patients received therapeutic benefit. The AUCs in the training set were all achieved at 0.926 (**Figure 5F**). Thus, the ICSscore value may be used as a predictive biomarker for immunotherapeutic benefit in melanoma.

DISCUSSION

A large number of studies have demonstrated that the TILs are associated with tumor progression and patient prognosis (Zheng et al., 2017; Ding et al., 2018; Lu et al., 2019). In the present study, on the basis of a comprehensive collection of marker genes, 182 ICSs associated with TIME were evaluated and applied. Here, an ICSscore formulated by the four-best prognosis-related ICS was constructed, which was validated successfully to predict prognosis and therapeutic benefit in HCC. Indeed, the four ICSs have significant associations to the tumor immune system, and a dozen marker genes attached to the four signatures have been reported to predict prognosis. For example, CHANG_CORE_SERUM_RESPONSE_UP was reported to correlate with wound healing, with elevated expression of angiogenic genes, a high proliferation rate, and a Th2 cell bias to the adaptive immune infiltrate. TREM1_data was marked by the only gene TREM1, which triggers phagocyte secretion of pro-inflammatory chemokines and cytokines. However, the specific biological roles of these ICSs remain to be further explored. In particular, the fitted ICSscore was found to be positively correlated with the risk level of HCC patients, but

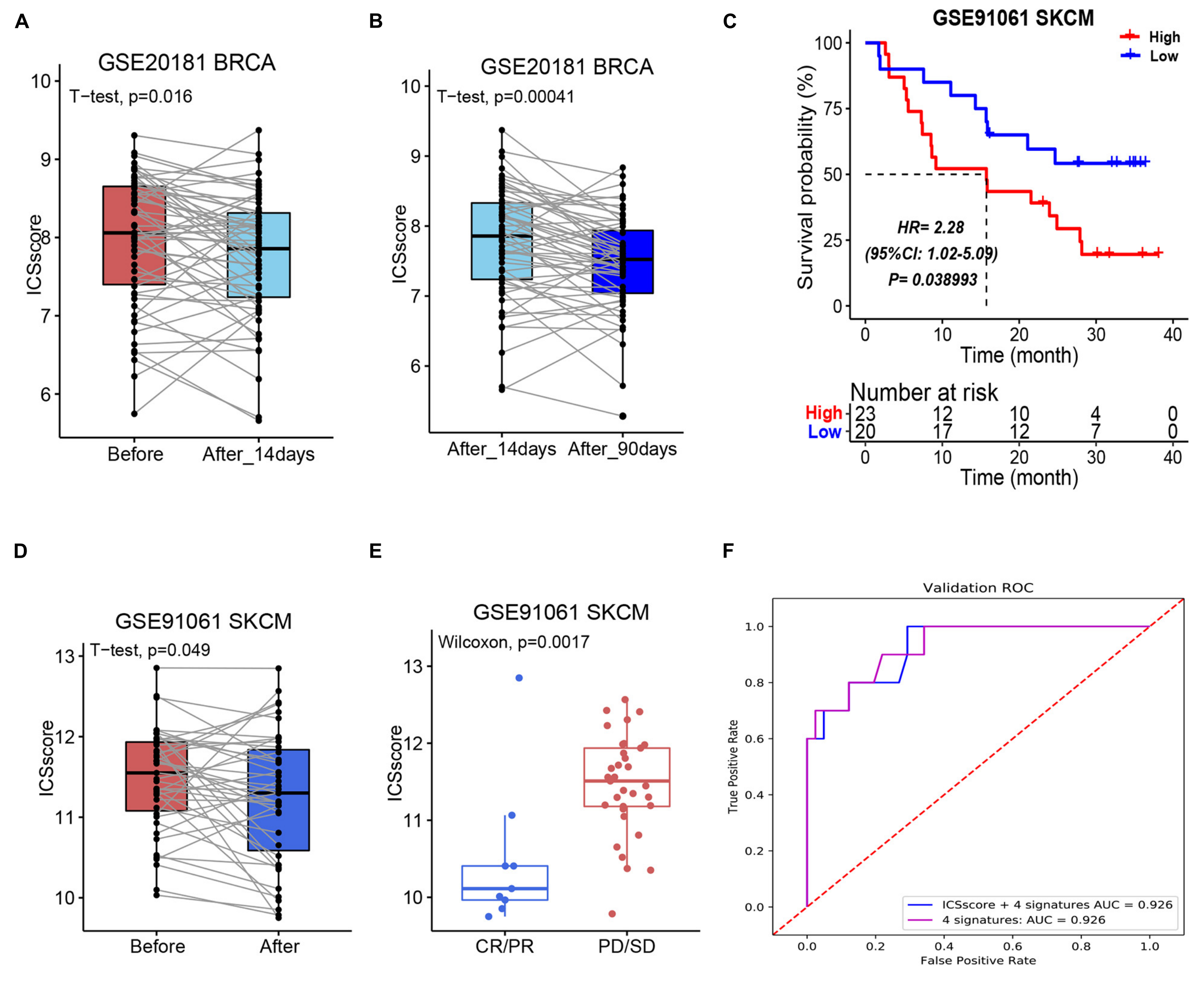


FIGURE 5 | Evaluation and prediction of chemotherapy and immunotherapy benefit in other tumors by ICSscore. **(A,B)** Pairwise comparison of the ICSscore in patients before and after chemotherapy in the GSE20181 BRCA cohort. **(C)** Kaplan–Meier curves of overall survival according to low- and high-ICSscore groups in the GSE91061 SKCM cohort. **(D)** Pairwise comparison of the ICSscore in patients before and after immunotherapy in the GSE91061 SKCM cohort. **(E)** Boxplot illustrates the ICSscore of patients with immunotherapy response in the GSE91061 SKCM cohort. **(F)** ROC curve of the XGBoost algorithm for predicting the therapeutic effects in the GSE91061 SKCM cohort.

negatively correlated with the therapeutic efficacy. That is to say, the fitted ICSscore not only can be used to predict prognosis, but also can be used as an effective biomarker to evaluate therapeutic benefit and monitor treatment efficacy.

Sorafenib has been considered the standard of care for patients with advanced unresectable HCC since 2007 (Abdelgalil et al., 2019). It is an important step to detect patients who would potentially benefit from sorafenib treatment. Here, we proved that the ICSscores were significantly reduced in sorafenib-responding HCC patients, indicating that the ICSscore may be a biomarker for predicting the response to sorafenib in HCC patients. Moreover, chemotherapy is one of the most important treatment modalities for advanced HCC. Significantly decreased ICSscores were observed in chemotherapy-responding HCC patients, indicating that the ICSscore can also be used as a marker for predicting the response to chemotherapy in HCC patients. Even so, due to the limitations of therapeutic datasets with regard to HCC, more real-world datasets are needed to further verify our findings and improve the ICSscore, especially those datasets using different treatments, such as immunotherapy. Similarly, gradually decreased ICSscore values were observed in breast cancer patients receiving chemotherapy for 14 days and 90 days, and significantly declined ICSscore values were found in melanoma patients with partial or complete remission after immunotherapy. These results imply that the ICSscore evaluation may be applied in pan cancer therapy supervision.

In recent years, immunotherapy exhibited promising therapeutic effects for advanced HCC, although only a few patients benefited from immunotherapy (Johnston and Khakoo, 2019; Riley et al., 2019; Zongyi and Xiaowu, 2020). Further research is needed to select effective biomarkers for patients who might benefit from immunotherapy. To our pleasure, the ICSscore evaluation may be used as a biomarker to distinguish patients who would respond to immunotherapy.

There were also some limitations in this study. Firstly, given that the large number of HCC patients used in this study came from different platforms, there may be significant batch effects in our cohort. Secondly, a series of ICSs were marked here, but only a few were used to construct the ICSscore. Thirdly, due to the limitation of datasets with treatment information, it is necessary to further testify and optimize ICSscore as a marker for immunotherapy in HCC, and even a broad spectrum of pan cancer.

CONCLUSION

Overall, we simplified the tedious ICSs to develop ICSscore, which can be applied successfully in prognostic stratification and therapeutic evaluation in HCC. Also, in melanoma and breast cancer, the unified ICSscore was validated to distinguish the samples with therapeutic benefits. This study provides a novel insight into the prognosis and therapeutic efficacy of ICS. ICSscore may be a potential marker for therapeutic efficacy in HCC, and even a broad spectrum of pan cancer.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

All procedures were in accordance with the ethical standards of the institutional and national committee on human experimentation and with the Helsinki Declaration (revised in 2013). There was no interaction with patients directly, as we acquired data from online public datasets.

AUTHOR CONTRIBUTIONS

LX, YL, LFX, and XJ conceived and designed this study. LFX and XJ were responsible for dataset collection, bioinformatics analysis, wrote the draft manuscript, and results interpretation. ZL, JZ, and SZ contributed to data analysis and discussion. LX revised the manuscript. The author(s) read and approved the final manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the National Natural Science Foundation of China (No. 31870829), the Shanghai Municipal Health Commission, and the Collaborative Innovation Cluster Project (No. 2019CXJQ02).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.741226/full#supplementary-material

Supplementary Figure S1 | Feature selection in TCGA-LIHC cohort. **(A)** The trend of the error rate changes with the depth of the treeand the variable importance ranking from random survival forest. **(B)** Theforest plot of the associations between the four selectedimmune cell signatures and overall survival in the TCGA cohort. The HR, 95% CI, and *p*-value weredetermined by multivariate Cox regression analysis.

Supplementary Figure S2 | The forest plot shows the association between ICSscore and overall survival in the subgroups of TCGA cohort.

Supplementary Figure S3 | Nomogram and Calibration developed in the TCGA-LIHC cohort. **(A)** A nomogram used to predict 1-year, 3-year, and 5-year overall survival probabilities. **(B)** Calibration curve of the overall survival of HCC patients for 1-, 3-, and 5-years.

Supplementary Figure S4 | The forest plot shows the association between ICSscore and overall survival in the subgroups of ICGC HCC cohort.

Supplementary Figure S5 | Nomogram and calibration developed in the ICGC HCC cohort. **(A)** A nomogram used to predict 1-year, 3-year, and 5-year overall survival probabilities. **(B)** Calibration curve of the overall survival of HCC patients for 1-, 3-, and 5-years.

Supplementary Figure S6 | The GO enrichment analysis based on the differentially expressed genes between high-risk and low-risk groups.

Supplementary Table S1 | The clinical characteristics of liver cancer patients in the three cohorts.

Supplementary Table S2 | 184 immune cell signatures and their corresponding genes.

Supplementary Table S3 | The significant immune cell signatures by performing univariate Cox regression analyses.

REFERENCES

Abdelgalil, A. A., Alkahtani, H. M., and Al-Jenoobi, F. I. (2019). Sorafenib. *Profiles Drug Subst. Excip. Relat. Methodol.* 44, 239–266. doi: 10.1016/bs.podrm.2018.11.003

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17:218. doi: 10.1186/s13059-016-1070-5

Bhattacharya, S., Andorf, S., Gomes, L., Dunn, P., Schaefer, H., Pontius, J., et al. (2014). ImmPort: disseminating data to the public for the future of immunology. *Immunol. Res.* 58, 234–239. doi: 10.1007/s12026-014-8516-1

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39, 782–795. doi: 10.1016/j.immuni.2013.10.003

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Bruix, J., Qin, S., Merle, P., Granito, A., Huang, Y. H., Bodoky, G., et al. (2017). Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* 389, 56–66. doi: 10.1016/s0140-6736(16)32453-9

Bruix, J., and Sherman, M. (2011). Management of hepatocellular carcinoma: an update. *Hepatology* 53, 1020–1022. doi: 10.1002/hep.24199

Chen, Y., Zhao, J., Jiao, Z., Wang, W., Wang, D., Yu, X., et al. (2018). SKA1 overexpression is associated with poor prognosis in hepatocellular carcinoma. *BMC Cancer* 18:1240. doi: 10.1186/s12885-018-5119-6

Ding, W., Xu, X., Qian, Y., Xue, W., Wang, Y., Du, J., et al. (2018). Prognostic value of tumor-infiltrating lymphocytes in hepatocellular carcinoma: a meta-analysis. *Medicine* 97:e13301. doi: 10.1097/md.0000000000013301

Frau, M., Ladu, S., Calvisi, D. F., Simile, M. M., Bonelli, P., Daino, L., et al. (2011). Mybl2 expression is under genetic control and contributes to determine a hepatocellular carcinoma susceptible phenotype. *J. Hepatol.* 55, 111–119. doi: 10.1016/j.jhep.2010.10.031

Fu, Y., Liu, S., Zeng, S., and Shen, H. (2019). From bench to bed: the tumor immune microenvironment and current immunotherapeutic strategies for hepatocellular carcinoma. *J. Exp. Clin. Cancer Res.* 38:396. doi: 10.1186/s13046-019-1396-4

Gabrielson, A., Wu, Y., Wang, H., Jiang, J., Kallakury, B., Gatalica, Z., et al. (2016). Intratumoral CD3 and CD8 T-cell densities associated with relapse-free survival in HCC. *Cancer Immunol. Res.* 4, 419–430. doi: 10.1158/2326-6066.Cir-15-0110

Garnelo, M., Tan, A., Her, Z., Yeong, J., Lim, C. J., Chen, J., et al. (2017). Interaction between tumour-infiltrating B cells and T cells controls the progression of hepatocellular carcinoma. *Gut* 66, 342–351. doi: 10.1136/gutjnl-2015-310814

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 14:7. doi: 10.1186/1471-2105-14-7

Heimbach, J. K., Kulik, L. M., Finn, R. S., Sirlin, C. B., Abecassis, M. M., Roberts, L. R., et al. (2018). AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology* 67, 358–380. doi: 10.1002/hep.29086

Hu, G., and Wang, S. (2017). Tumor-infiltrating CD45RO(+) memory T lymphocytes predict favorable clinical outcome in solid tumors. *Sci. Rep.* 7:10376. doi: 10.1038/s41598-017-11122-2

Ishwaran, H. (2015). The effect of splitting on random forests. *Mach. Learn.* 99, 75–118. doi: 10.1007/s10994-014-5451-2

Jiang, Y. Q., Cao, S. E., Cao, S., Chen, J. N., Wang, G. Y., Shi, W. Q., et al. (2021). Preoperative identification of microvascular invasion in hepatocellular carcinoma by XGBoost and deep learning. *J. Cancer Res. Clin. Oncol.* 147, 821–833. doi: 10.1007/s00432-020-03366-9

Johnston, M. P., and Khakoo, S. I. (2019). Immunotherapy for hepatocellular carcinoma: current and future. *World J. Gastroenterol.* 25, 2977–2989. doi: 10.3748/wjg.v25.i24.2977

Kuhlmann, J. B., and Blum, H. E. (2013). Locoregional therapy for cholangiocarcinoma. *Curr. Opin. Gastroenterol.* 29, 324–328. doi: 10.1097/MOG.0b013e32835d9dea

Kurebayashi, Y., Ojima, H., Tsujikawa, H., Kubota, N., Maehara, J., Abe, Y., et al. (2018). Landscape of immune microenvironment in hepatocellular carcinoma and its additional impact on histological and molecular classification. *Hepatology* 68, 1025–1041. doi: 10.1002/hep.29904

Lian, Q., Wang, S., Zhang, G., Wang, D., Luo, G., Tang, J., et al. (2018). HCCDB: a database of hepatocellular carcinoma expression atlas. *Genomics Proteomics Bioinformatics* 16, 269–275. doi: 10.1016/j.gpb.2018.07.003

Liu, C., Seeram, N. P., and Ma, H. (2021). Small molecule inhibitors against PD-1/PD-L1 immune checkpoints and current methodologies for their development: a review. *Cancer Cell Int.* 21:239. doi: 10.1186/s12935-021-01946-4

Liu, P., Wei, J., Mao, F., Xin, Z., Duan, H., Du, Y., et al. (2021). Establishment of a prognostic model for hepatocellular carcinoma based on endoplasmic reticulum stress-related gene analysis. *Front. Oncol.* 11:641487. doi: 10.3389/fonc.2021.641487

Liu, P. H., Hsu, C. Y., Hsia, C. Y., Lee, Y. H., Su, C. W., Huang, Y. H., et al. (2016). Prognosis of hepatocellular carcinoma: assessment of eleven staging systems. *J. Hepatol.* 64, 601–608. doi: 10.1016/j.jhep.2015.10.029

Llovet, J. M., Zucman-Rossi, J., Pikarsky, E., Sangro, B., Schwartz, M., Sherman, M., et al. (2016). Hepatocellular carcinoma. *Nat. Rev. Dis. Primers* 2:16018. doi: 10.1038/nrdp.2016.18

Lu, C., Rong, D., Zhang, B., Zheng, W., Wang, X., Chen, Z., et al. (2019). Current perspectives on the immunosuppressive tumor microenvironment in hepatocellular carcinoma: challenges and opportunities. *Mol. Cancer* 18:130. doi: 10.1186/s12943-019-1047-6

Miao, Y. R., Zhang, Q., Lei, Q., Luo, M., Xie, G. Y., Wang, H., et al. (2020). ImmuCellAI: a unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy. *Adv. Sci.* 7:1902880. doi: 10.1002/advs.201902880

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457. doi: 10.1038/nmeth.3337

Nirmal, A. J., Regan, T., Shih, B. B., Hume, D. A., Sims, A. H., and Freeman, T. C. (2018). Immune cell gene signatures for profiling the microenvironment of solid tumors. *Cancer Immunol. Res.* 6, 1388–1400. doi: 10.1158/2326-6066.Cir-18-0342

Nishida, N., and Kudo, M. (2017). Immunological microenvironment of hepatocellular carcinoma and its clinical implication. *Oncology* 92(Suppl. 1), 40–49. doi: 10.1159/000451015

Olsen, S. K., Brown, R. S., and Siegel, A. B. (2010). Hepatocellular carcinoma: review of current treatment with a focus on targeted molecular therapies. *Therap. Adv. Gastroenterol.* 3, 55–66. doi: 10.1177/1756283x09346669

Park, S., Choi, S., Cho, Y. A., Sinn, D. H., Kim, J. M., Park, C. K., et al. (2020). Evaluation of the American Joint Committee on Cancer (AJCC) 8th Edition staging system for hepatocellular carcinoma in 1,008 patients with curative resection. *Cancer Res. Treat.* 52, 1145–1152. doi: 10.4143/crt.2020.208

Riaz, N., Havel, J. J., Makarov, V., Desrichard, A., Urba, W. J., Sims, J. S., et al. (2017). Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell* 171, 934–949.e16. doi: 10.1016/j.cell.2017.09.028

Riley, R. S., June, C. H., Langer, R., and Mitchell, M. J. (2019). Delivery technologies for cancer immunotherapy. *Nat. Rev. Drug Discov.* 18, 175–196. doi: 10.1038/s41573-018-0006-z

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007

Tang, X., Shu, Z., Zhang, W., Cheng, L., Yu, J., Zhang, M., et al. (2019). Clinical significance of the immune cell landscape in hepatocellular carcinoma patients with different degrees of fibrosis. *Ann. Transl. Med.* 7:528. doi: 10.21037/atm.2019.09.122

Wang, S., Xiong, Y., Zhang, Q., Su, D., Yu, C., Cao, Y., et al. (2020). Clinical significance and immunogenomic landscape analyses of the immune cell signature based prognostic model for patients with breast cancer. *Brief. Bioinform.* 22:bbaa311. doi: 10.1093/bib/bbaa311

Wang, Y., Xie, Y., Ma, J., Wang, Y., and Gong, R. (2020). Development and validation of a prognostic and immunotherapeutically relevant model in hepatocellular carcinoma. *Ann. Transl. Med.* 8:1177. doi: 10.21037/atm-20-6112

Wang, Z., Wang, Y., Yang, T., Xing, H., Wang, Y., Gao, L., et al. (2021). Machine learning revealed stemness features and a novel stemness-based classification with appealing implications in discriminating the prognosis, immunotherapy and temozolomide responses of 906 glioblastoma patients. *Brief. Bioinform.* 22:bbab032. doi: 10.1093/bib/bbab032

Wolf, D. M., Lenburg, M. E., Yau, C., Boudreau, A., and van ‘t Veer, L. J. (2014). Gene co-expression modules as clinically relevant hallmarks of breast cancer diversity. *PLoS One* 9:e88309. doi: 10.1371/journal.pone.0088309

Zhang, B., Tang, B., Gao, J., Li, J., Kong, L., and Qin, L. (2020). A hypoxia-related signature for clinically predicting diagnosis, prognosis and immune microenvironment of hepatocellular carcinoma patients. *J. Transl. Med.* 18:342. doi: 10.1186/s12967-020-02492-9

Zhang, Q., He, Y., Luo, N., Patel, S. J., Han, Y., Gao, R., et al. (2019). Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell* 179, 829–845.e20. doi: 10.1016/j.cell.2019.10.003

Zheng, C., Zheng, L., Yoo, J. K., Guo, H., Zhang, Y., Guo, X., et al. (2017). Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell* 169, 1342–1356.e16. doi: 10.1016/j.cell.2017.05.035

Zhou, Z., Zhou, Z., Huang, Z., He, S., and Chen, S. (2020). Histone-fold centromere protein W (CENP-W) is associated with the biological behavior of hepatocellular carcinoma cells. *Bioengineered* 11, 729–742. doi: 10.1080/21655979.2020.1787776

Zongyi, Y., and Xiaowu, L. (2020). Immunotherapy for hepatocellular carcinoma. *Cancer Lett.* 470, 8–17. doi: 10.1016/j.canlet.2019.12.002

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 02 December 2021
doi: 10.3389/fgene.2021.770857

Intra-Tumoral Expression of SLC7A11 Is Associated with Immune Microenvironment, Drug Resistance, and Prognosis in Cancers: A Pan-Cancer Analysis

Jiajun He, Hongjian Ding, Huaqing Li, Zhiyu Pan and Qian Chen**

Department of General Surgery, Minhang Hospital, Fudan University, Shanghai, China

While many anti-cancer modalities have shown potent efficacy in clinical practices, cancer prevention, timely detection, and effective treatment are still challenging. As a newly recognized iron-dependent cell death mechanism characterized by excessive generation of lipid peroxidation, ferroptosis is regarded as a potent weapon in clearing cancer cells. The cystine/glutamate antiporter solute carrier family 7 member 11 (SLC7A11) is the core target for ferroptosis regulation, the overexpression of which dictates downregulated sensitivity to ferroptosis in cancer cells. Hence, we elaborated the pan-cancer level bioinformatic study and systematically elucidated the role of intra-tumoral expression of SLC7A11 in the survival of cancer patients and potential immunotherapeutic response. Specifically, 25/27 (92.6%) cancers were featured with upregulated SLC7A11 expression, where SLC7A11 overexpression is a risk factor for worse overall survival in 8 cancers. We also validated SLC7A11 expression in multiple pancreatic cancer cell lines *in vitro* and found that it was upregulated in most pancreatic cancer cell lines ($p < 0.05$). Single-cell sequencing method revealed the SLC7A11 was majorly expressed in cancer cells and mononuclear cells. To further explore the function of SLC7A11 in cancer progression, we analyzed the influence on cell proliferation after the knockdown or knockout of SLC7A11 by either CRISPR or RNAi methods. Besides, the association between SLC7A11 and drug resistance was characterized using bioinformatic approaches as well. We also analyzed the association between the expression of SLC7A11 in multi-omics level and the intra-tumoral infiltration of immune cells based on cell annotation algorithms. Moreover, the relationship between SLC7A11 and the expression of MHC, immune stimulators, immune inhibitors as well as the response to immunotherapy was investigated. In addition, the SLC7A11 expression in colon adenocarcinoma, uterine corpus endometrial carcinoma, and stomach adenocarcinoma (STAD) is also positively associated with microsatellite instability and that in head and neck squamous cell carcinoma, STAD, and prostate adenocarcinoma is positively associated with neoantigen level, which further revealed the potential relationship between SLC7A11 and immunotherapeutic response.

Keywords: ferroptosis, SLC7A11, pan-cancer, immune microenvironment, drug resistance

Edited by:
Luis Zapata,
Institute of Cancer Research (ICR),
United Kingdom

Reviewed by:
Shankar Suman,
The Ohio State University,
United States
Juan Wang,
Inner Mongolia University, China

***Correspondence:**
Qian Chen
chenqian@fudan.edu.cn
Jiajun He
doctorhejiajun@163.com

Specialty section:
This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: *05 September 2021*
Accepted: *01 November 2021*
Published: *02 December 2021*

Citation:
He J, Ding H, Li H, Pan Z and Chen Q (2021) Intra-Tumoral Expression of SLC7A11 Is Associated with Immune Microenvironment, Drug Resistance, and Prognosis in Cancers: A Pan-Cancer Analysis.
Front. Genet. 12:770857.
doi: 10.3389/fgene.2021.770857

INTRODUCTION

Cancer is the leading cause of human deaths in the world, which produces serious economic burdens both in developed and developing countries (Wu et al., 2019). Although many anti-cancer modalities, such as neoadjuvant chemotherapy and immunotherapy, have shown potent efficacy in clinical practice, cancer prevention, early detection, and effective treatment are still challenging in most cases (O'Donnell et al., 2019; Yang, 2015; Burotto et al., 2019). Hence, finding a more effective strategy to treat cancers is necessary and urgent.

Ferroptosis, a newly recognized iron-dependent cell death mechanism, is characterized by excessive generation of lipid peroxidation (Tang et al., 2020). Recently, an increasing number of studies reported that ferroptosis is involved in many pathophysiological conditions, including cardiovascular diseases, neurodegenerative diseases, and cancers (Jeong et al., 2017; Liang and Zhang, 2019; Mou et al., 2019). Many studies have reported that ferroptosis-induced cell death is an effective approach in killing cancer cells through reactive oxygen species (ROS) accumulation in cells, although its clinical benefits still need clinical trials for verification (Friedmann Angeli et al., 2019; Hassannia et al., 2019). An *in vivo* study showed that inducing tumor-selective ferroptosis via deletion of a system xC-subunit (SLC7A11) was dramatically contributing to the inhibition of the growth of pancreatic ductal adenocarcinoma, which is one of the most lethal solid organ malignancies (Bai et al., 2018). In addition, Wang et al. demonstrated that PD-1-based immunotherapy combined with ferroptosis induction has a synergistic effect compared with single treatment. Given the low response rate of immunotherapy, ferroptosis induction may be a potent adjuvant modality in anti-cancer immunotherapy (Wang et al., 2019).

As a nutrient transporter frequently overexpressed in human malignancies, SLC7A11 is the cystine/glutamate antiporter solute carrier family 7 member 11 (SLC7A11; also known as xCT) (Sehm et al., 2016; Liu et al., 2019). SLC7A11 could promote cystine uptake and glutathione biosynthesis, leading to protection from oxidative stress and ferroptotic cell death (Sehm et al., 2016; Liu et al., 2019). Depleting SLC7A11 dramatically decreased glutathione concentrations and triggered ferroptosis. A study using genetically engineered mice with SLC7A11 knockout revealed tumor-selective ferroptosis and inhibited the growth of pancreatic cancer (Badgley et al., 2020). In view of its important role in ferroptosis execution, the major reagents that induce ferroptosis are targeted at SLC7A11 like erastin (Bai et al., 2018; Shibata et al., 2019). Several studies have also revealed that SLC7A11 plays vital roles in glutamine metabolism and regulates the glucose and glutamine dependency of cancer cells (Shin et al., 2017). Interestingly, the components in tumor microenvironment could also promote or restrain tumor ferroptotic cell death by influencing the SLC7A11 expression level (Wang et al., 2019; Li et al., 2020; Zhang et al., 2020). CD8$^+$ T cells induced ferroptosis in tumor cells through IFN-y-dependent SLC7A11 downregulation (Wang et al., 2019). On the contrary, cancer-associated fibroblasts suppress ferroptosis and promote acquired chemoresistance in gastric cancer through secreting miR-522 (Zhang et al., 2020). Hence, it is significant to investigate the role of SLC7A11 in cancer development. Here, we conducted a bioinformatic study to systematically explore whether the intra-tumoral expression of SLC7A11 is associated with cancer patients' prognosis and response to immunotherapy.

MATERIALS AND METHODS

The Source of Transcriptome and Clinical Data

The transcriptome data of 33 cancers, including adrenocortical carcinoma (ACC), bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), cervical squamous cell carcinoma and endocervical adenocarcinoma (CSEA), lymphoid neoplasm diffuse large B-cell lymphoma (DLBC), esophageal carcinoma (ESCA), glioblastoma multiforme (GBM), head-and-neck squamous cell carcinoma (HNSC), kidney chromophore (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), acute myeloid leukemia (LAML), brain lower grade glioma (LGG), liver hepatocellular carcinoma (LIHC), lung squamous cell carcinoma (LUSC), lung adenocarcinoma (LUAD), mesothelioma (MESO), ovarian serous cystadenocarcinoma (OV), pheochromocytoma and paraganglioma (PCPG), pancreatic adenocarcinoma (PAAD), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), sarcoma (SARC), skin cutaneous melanoma (SKCM), testicular germ cell tumor (TGCT), thyroid carcinoma (THCA), stomach adenocarcinoma (STAD), thymoma (THYM), uterine corpus endometrial carcinoma (UCEC), uterine carcinosarcoma (UCS), and uveal melanoma (UVM), were downloaded from The Cancer Genome Atlas (TCGA). We merged the transcriptome data of normal pancreas in The Genotype Tissue Expression (GTEx), which is a comprehensive public resource to study tissue-specific gene expression, with TCGA cohort given that the latter lacks normal samples for control. Besides, the expression of SLC7A11 is also evaluated in distinct cancer cell lines through Cancer Cell Line Encyclopedia (CCLE) (https://portals.broadinstitute.org/ccle/data). Fragments per kilobase million (FPKM) was selected as the data format for following calculation. The patients' clinical information, including overall survival (OS), disease-specific survival (DSS), disease-free interval (DFI), and progression-free interval (PFI), was also downloaded from TCGA and merged with transcriptome data.

Bioinformatic and Statistical Analysis

A univariate Cox regression analysis was applied to identify the association between SLC7A11 expression and OS, DSS, DFI, and PFI across 33 cancers. Hazard ratio (HR) was used to evaluate the magnitude of association with R package "survival" (version 3.1-8). Kaplan-Meier survival curve was depicted to visualize those associations with statistical significance.

Optimum cutoff value was determined dependent of the largest Youden index. R Package "Estimate" (version 1.013) was used to estimate the proportion of immune and stromal cells in malignant tumor tissues from transcriptome data.

Specifically, we introduced "immune score" to assess the proportion of immune cells and "stromal score" to assess the proportion of stromal cells. Estimate score is equal to the sum of immune and stromal scores. Pearson correlation coefficient (r) was used to evaluate the strength of association between SLC7A11 expression and immune/stromal/estimate scores.

The expression level of 47 immune checkpoint genes was extracted from transcriptome data of each cancer. The co-expression association was also calculated using Pearson correlation coefficient and visualized as a heatmap. The infiltration of six common immune cells, $CD8^+$ T cells, $CD4^+$ T cells, B cells, macrophages, neutrophils, and dendritic cells, was evaluated using Tumor Immune Estimation Resource (TIMER) database. Tumor mutation burden (TMB) is defined as the total number of somatic gene coding errors, base substitutions, and gene insertions or deletions detected per million bases. We calculated the TMB of each cancer sample based on the exome sequencing data from TCGA (VarScan2). MSI (microsatellite instability) referred to the molecular fingerprint of a deficient mismatch repair system. We referred to previous studies for summarized MSI data across distinct cancers (Hause et al., 2016; Yang et al., 2019). Neoantigens are encoded by mutated genes of tumor cells, which are mainly new abnormal proteins produced by gene point mutation, deletion mutation, and gene fusion that are different from proteins expressed by normal cells. Neoantigen level was obtained from a previously published study (Rooney et al., 2015). The correlation between TMB/MSI/neoantigen and SLC7A11 expression was calculated using Pearson correlation coefficient and visualized as radar plots. Similarly, the association between SLC7A11 expression and the expression level of DNA repair-related regulators 17 and methyltransferases were also evaluated by Pearson correlation coefficient.

Gene set enrichment analysis (GSEA) was performed to identify which pathways are varied along with the differential expression of SLC7A11. The top five enriched pathways were showcased according to the ranking of normalized enrichment score. The results in this study were seen statistically significant when p value is less than 0.05.

To analyze the association between SLC7A11 expression and drug resistance in pan-cancer landscape, we first used RNAactDrug database (http://bio-bigdata.hrbmu.edu.cn/RNAactDrug). A total of 562 FDA-approved drugs were analyzed using three common methods, including CellMiner, GDSC, and CCLE.

To present the single-cell transcriptomic expression of SLC7A11, we applied TISCH method to analyze the expression pattern of SLC7A11 in different types of cells in tumor microenvironment (http://tisch.comp-genomics.org/).

We also analyzed the correlation between the methylation, mRNA abundance, and copy number of SLC7A11 and immune infiltration, MHC, and immune stimulators and inhibitors in multiple cancers (http://cis.hku.hk/TISIDB/). To further study the relationship between SLC7A11 expression and T-cell dysfunction and immunotherapeutic response, we turn to TIDE algorithm and performed relevant analysis (http://tide.dfci.harvard.edu/).

In Vitro Validation

The human pancreatic cancer cell line HPDE, BxPC-3, AsPC-1, Capan-1, Panc-1, SW1990, and Mia-Paca2 were obtained from the American Type Culture Collection. Capan-1 cells were cultured in Iscove's modified Dulbecco's medium with 10% fetal bovine serum. Panc-1, SW1990, and Mia-Paca2 were cultured in Dulbecco's modified eagle medium (DMEM). BxPC-3 and AsPC-1 were cultured in Gibco Roswell Park Memorial Institute (RPMI) 1640 Medium. Then, RNA was extracted from cell lines using SteadyPure Universal RNA Extraction Kit (AG21017, China). Quantitative real-time PCR was performed using SYBR green (Qiagen, USA).

To analyze the influence on cell proliferation following by SLC7A11 knockout, we applied DepMap tool (https://depmap.org/) to assess the CERES value for SLC7A11 in cell lines in pan-cancer level. Notably, a smaller CERES value reflected the stronger ability to promote cell proliferation *in vitro*.

Small Interfering RNA-Mediated Knockdown and CCK-8 Assay

The cells (2×10^5) were seeded in a six-well culture plate. After resting overnight, Lipofectamine 3000 reagent (Invitrogen, Thermo Fisher Scientific, Carlsbad, CA, USA) and the indicated concentration of siRNA were added in serum-free DMEM. The complex was added to antibiotic-free medium. After 48 h, the cells were collected for further experiments. Cell relative viability was determined daily using CCK-8 (Beyotime: C0037) based on the manufacturer's instructions.

RESULTS

The Differential Expression of SLC7A11 in Distinct Cancer Cell Lines, Normal and Tumor Samples Within Bulk or Single-Cell Transcriptomic Landscape

The expression level of SLC7A11 is highly inconsistent across 31 normal tissues (Kruskal-Wallis test $p < 0.05$) (**Figure 1A**). Notably, SLC7A11 is highly expressed in bone marrow but rarely expressed in adipose tissue, adrenal gland, cervix uteri, kidney, liver, muscle, nerve, and uterus. This phenomenon suggested that not all organs are sensitive to ferroptosis due to the differential expression of SLC7A11. Because SLC7A11 has a low expression in normal tissues, ferroptosis is thought to occur in normal cells and could be a physiologic activity. Then, we further compared the expression level among cancer cell lines (**Figure 1B**). Interestingly, the landscape of SLC7A11 expression across cancers changed dramatically. Specifically, SLC7A11 has the highest expression level in pleural but rarely expressed in breast, hematopoietic/lymphoid, and soft tissues. Hence, SLC7A11 may experience a transcriptome remodeling during the carcinogenesis. To support our assumption, we further compared the differential expression of SLC7A11 between human normal and tumor tissues using TCGA data (**Figure 1C**). The results showed that SLC7A11 is

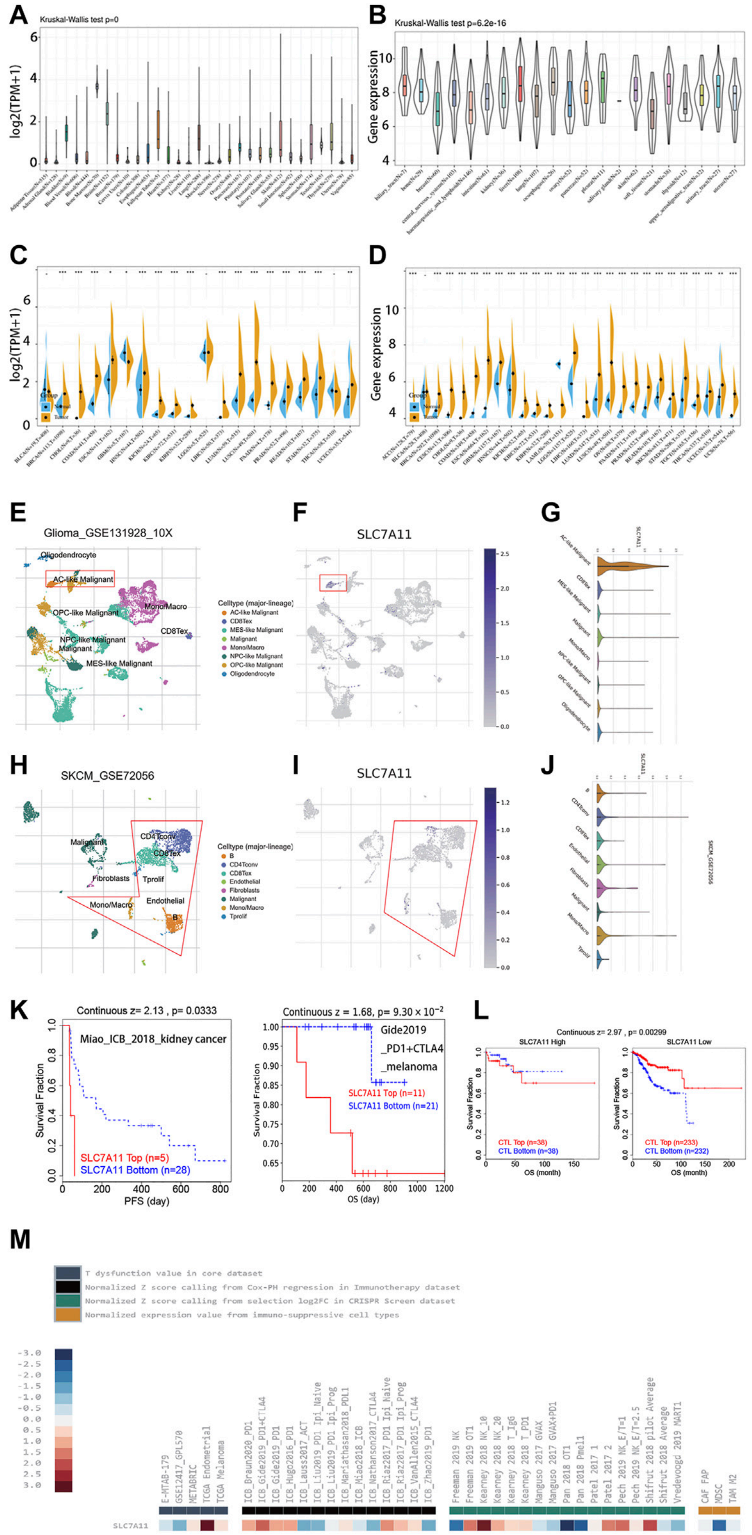


FIGURE 1 | The differential expression of SLC7A11 in distinct cancers. **(A)** The expression level of SLC7A11 in 33 normal organs. **(B)** The expression level of SLC7A11 in 21 cancer cell lines. **(C)** Investigation of the differential expression of SLC7A11 between cancers and normal tissues using TCGA data. **(D)** Investigation of the differential expression of SLC7A11 between cancers and normal tissues using the data from TCGA combined with GTEx. **(E–J)** Single-cell analysis and TIDE algorithms revealing the expression distribution and immunosuppressive characteristic of SLC7A11 in cell clusters. **(K)** Immunotherapy prolonged the survival expectancy only in patients with lower SLC7A11 expression. **(L)** SLC7A11 is associated with increased CTL cytotoxicity based on TIDE algorithm; CTL top referred to samples with higher CTL infiltration, *(Continued)*

FIGURE 1 | while CTL bottom means lower CTL infiltration. The cut-off value to distinguish high and low CTL infiltration is just dependent on the Cox-PH model embedded in TIDE algorithm. **(M)** Comprehensive presentation of the influence on T-dysfunction and immunotherapeutic response associated with SLC7A11 by TIDE algorithm.

overexpressed in 80% (16/20) cancers, which indicated that many cancers may shrink ferroptosis by upregulating SLC7A11 expression. Due to the insufficiency of normal samples in some tumor categories in TCGA, we further incorporated the transcriptome data from GTEx and re-evaluated the differential expression of SLC7A11 in 27 cancers (**Figure 1D**). The results noted that 92.6% cancers were featured with upregulated SLC7A11 expression, which further supported our hypothesis. We also validated the differential expression of SLC7A11 in distinct cell lines. Compared with normal pancreatic ductal cells, the expression of SLC7A11 was upregulated in most pancreatic cancer cell lines ($p < 0.05$; **Supplementary Figure S1**).

Single-cell transcriptomic analysis revealed the expression pattern of SLC7A11 in different cell types (**Figures 1E–J**; **Supplementary Figure S2**). First, SLC7A11 is rarely expressed in hematologic malignancies (AEL, AML, and ALL). In solid tumors, SLC7A11 was observed to express highly in the malignant cells of glioma, while for SKCM, SLC7A11 was highly expressed in immune cells and moderately expressed in malignant and stromal cells. For STAD and UCEC, SLC7A11 was more preferred to express in stromal cells. In addition, immunotherapy may only function in patients with lower SLC7A11 expression (**Figure 1K**). We also presented the comprehensive landscape for the relationship between SLC7A11 and T-dysfunction or immunotherapeutic response in the core cohorts in TIDE database (**Figures 1L,M**).

The Correlation Between Intra-Tumoral SLC7A11 Expression and Patients' Overall Survival, Disease-Specific Survival, Disease-Free Interval, and Progression-Free Interval

Given the obviously differential expression of SLC7A11 observed between tumor and normal samples, we further investigate whether SLC7A11 expression influences patients' prognosis. OS, DSS, DFI, and PFI were selected as the indicators to reflect patients' prognosis. Overexpression of SLC7A11 was identified as a risk factor for worse OS in eight cancers (ACC, BLCA, HNSC, KICH, KIRC, LGG, LIHC, and SKCM) (HR > 1, $p < 0.05$). On the contrary, SLC7A11 served as a protective factor for prolonged OS in OV and READ (HR < 1, $p < 0.05$) (**Figure 2A** and **Figure 2E**). In addition, SLC7A11 overexpression was an adverse factor for the DSS of patients with eight cancers (ACC, BLCA, KICH, KIRC, KIRP, LGG, PRAD, and SKCM) (HR > 1, $p < 0.05$); however, it was a favorable factor for better DSS in READ (HR < 1, $p < 0.05$) (**Figure 2B** and **Figure 2F**). SLC7A11 is negatively associated with longer DFI in ACC, LGG, and KIRP (HR > 1, $p < 0.05$) and better PFI in ACC, BLCA, HNSC, KICH, KIRC, KIRP, and LGG (HR > 1, $p < 0.05$) (**Figures 2C,G**, respectively). Overall, excessive expression of intra-tumoral SLC7A11 may be an unfavorable factor for patients' prognosis in several cancers. It is biologically plausible that tumor cells upregulate SLC7A11 expression to shirk ferroptosis execution and further undermine patients' survival.

SLC7A11 Barely Influenced Tumor Cell Proliferation *In Vitro*

Given that SLC7A11 was associated with the prognoses in multiple cancers, we studied whether SLC7A11 affected the proliferation of cancer cells. We analyzed the CRISPR- and RNAi-based data in DEPMAP, and the results derived from 989 to 709 cell lines, respectively, showed SLC7A11 knockout or knockdown barely compromised the proliferation ability of all kinds of tumor cells (**Figure 3A**), which suggested SLC7A11 is associated with the worse prognosis of cancer patients through other mechanisms, tumor microenvironment, for example. To further confirm this finding, we knocked down the expression of SLC7A11 *in vitro* in Panc-1 cell and found SLC7A11 knockdown did not affect the proliferation of pancreatic cancer (**Figures 3C,D**). We also identified the differentially expressed molecules after SLC7A11 knockdown or knockout, which could be structurally interacted molecules for SLC7A11 (**Supplementary Table S1**).

We further analyzed the association between SLC7A11 expression and drug sensitivity during cancer treatment using RNAactDrug database. Among 562 FDA-approved drugs, we found the expression of SLC7A11 in mRNA level was significantly positively correlated with six drugs (R > 0.4 and fdr<0.01), including selendale, gelcohol, n6-benzyladenosine-5′-phosphate, 6-benzylthioinosine, (E)-3-(3-((1-(4-fluorobenzyl)-1H-1,2,3-triazol-4-yl)met, kinetin riboside, adenosine, 8-chloro-, cyclic 3′,5′-(hydrogen phosphate), and (E)-1-(benzo[d] (Yang, 2015; Wu et al., 2019)dioxol-5-yl) (**Supplementary Table S2**). Besides, we investigated the relationship between SLC7A11 and drug resistance in TISIDB, the results of which showed small molecule riluzole, thimerosal, cystine, acetylcysteine, sulfasalazine, and glutamic acid (**Figure 3B**). Notably, SLC7A11 was a key regulator for the transportation of cystine and glutamic acid, which supported the results yield by analyzing TISIDB tool.

The Correlation Between SLC7A11 Expression and Immune Cell Infiltration, Immune Score, Stromal Score, and Estimate Score

Recently, a well-conducted study reported that $CD8^+$ T cells could induce ferroptosis in tumor cells via downregulating the SLC7A11 expression (Wang et al., 2019). The combined

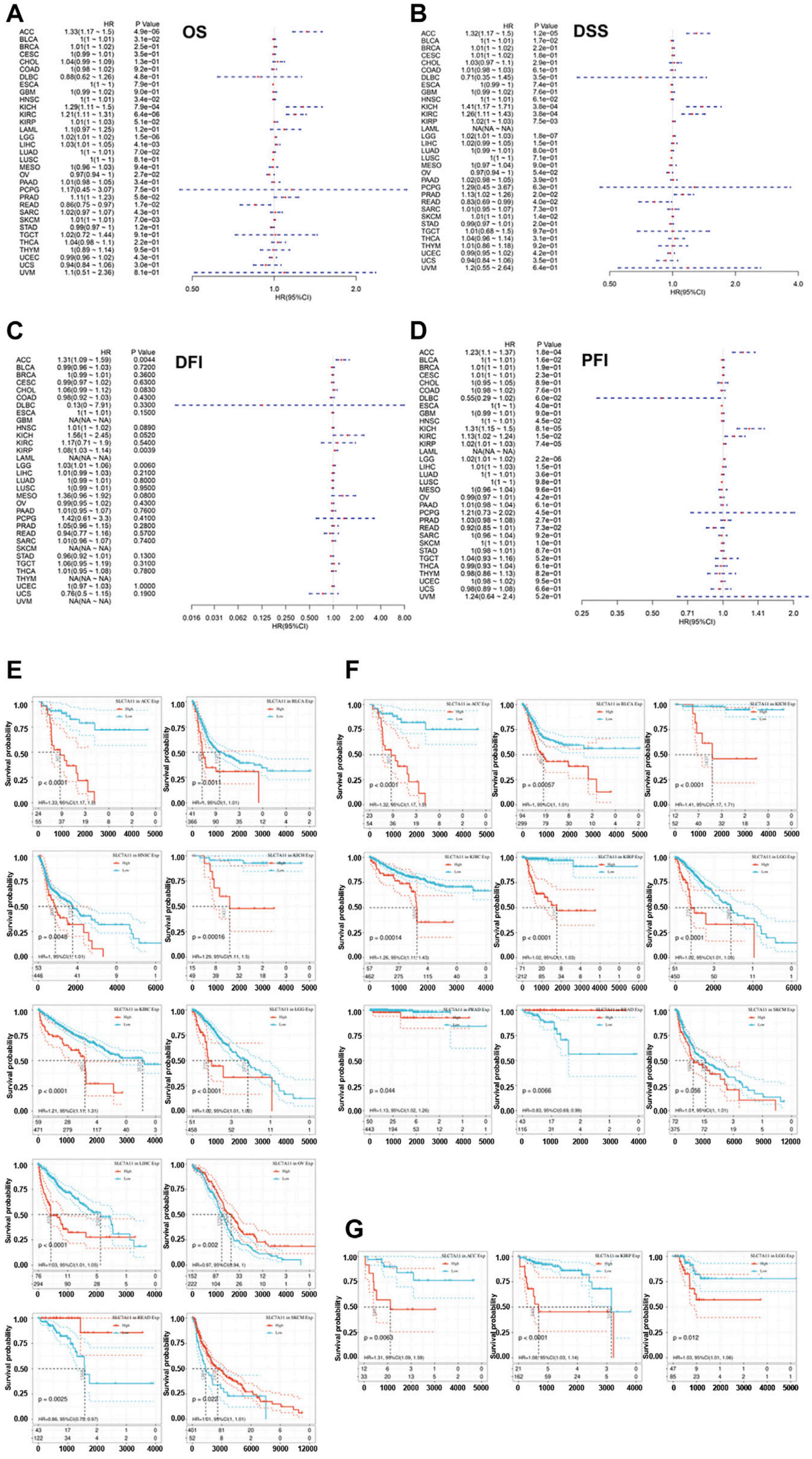


FIGURE 2 | Correlation of the expression of SLC7A11 with the prognosis of distinct cancers. **(A)** The association between SLC7A11 level and overall survival of cancers. **(B)** The association between SLC7A11 level and disease-specific survival of cancers. **(C)** The association between SLC7A11 level and disease-free interval of cancers. **(D)** The association between SLC7A11 level and progression-free interval of cancers. **(E–G)** The survival curve revealed that SLC7A11 is significantly associated with the prognosis of several cancers.

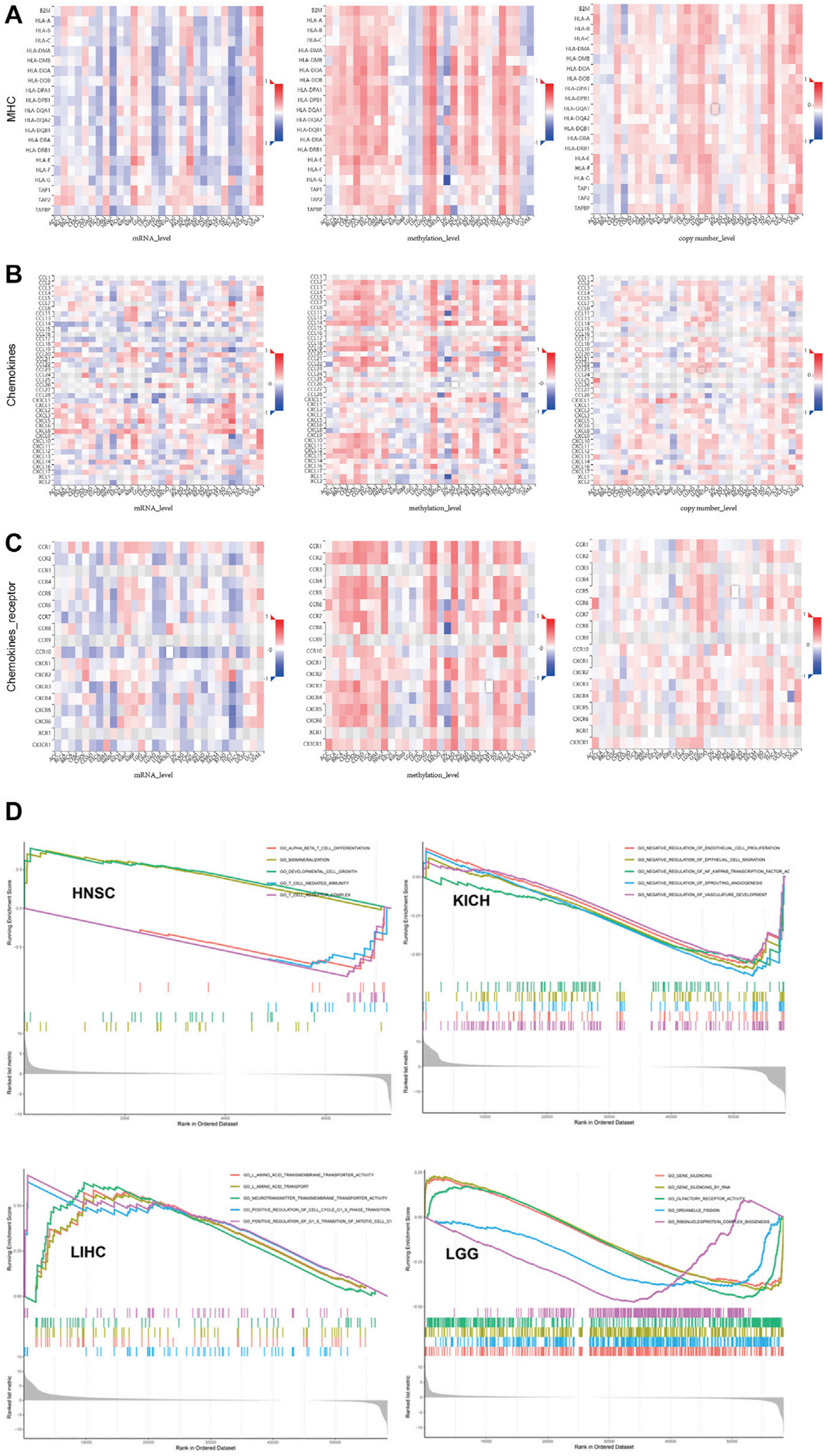
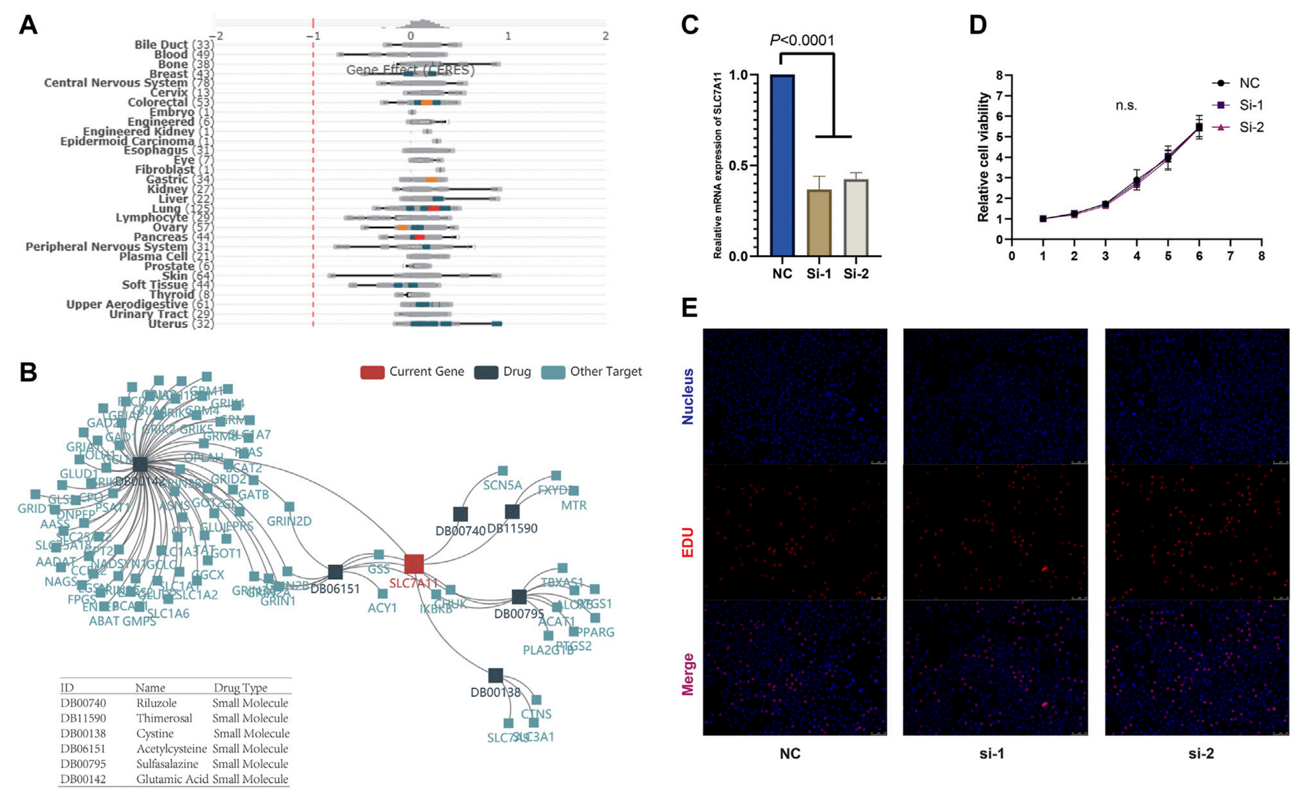


FIGURE 3 | Multi-omics data showed SLC7A11 expression was broadly associated with immune biomarkers but with tumor specificity. **(A)** MHC. **(B)** Chemokines. **(C)** Chemokine receptor. **(D)** GSEA to investigate which signal pathways were altered along with differential expression of SLC7A11 in cancers whose prognosis was associated with SLC7A11.



ID	Name	Drug Type
DB00740	Riluzole	Small Molecule
DB11590	Thimerosal	Small Molecule
DB00138	Cystine	Small Molecule
DB06151	Acetylcysteine	Small Molecule
DB00795	Sulfasalazine	Small Molecule
DB00142	Glutamic Acid	Small Molecule

FIGURE 4 | (A) SLC7A11 expression barely affects tumor cell proliferation *in vitro*. **(B)** SLC7A11 expression is associated with the sensitivity to multiple drugs. **(C)** Knockdown of SLC7A11. **(D, E)** CCK-8 and EDU analysis shows there are no obvious changes in cell proliferation after SLC7A11 knockdown in panc-1 cell.

induction of ferroptosis with immune checkpoint inhibitors has a synergistic effect in anti-tumor therapy. Hence, we speculated SLC7A11 is associated with the immune microenvironment of cancers. The association between SLC7A11 expression level and immune cell infiltration, immune score, stromal score, and estimate score was evaluated using Pearson correlation coefficient (**Figure 4A**). The results showed that SLC7A11 expression is negatively associated with CD8$^+$ T-cell infiltration in seven cancers (DLBC, ESCA, HNSC, LUAD, LUSC, TGCT, and THCA), suggesting the combination of ferroptosis induction and immunotherapy may be suitable in these cancers' treatment (**Supplementary Figure S3**). In addition, the association between immune score and SLC7A11 expression is dependent on cancer type. For PAAD, CESC, ESCA, HNSC, LUAD, LUSC, TGCT, and THCA, SLC7A11 expression is negatively associated with immune score. On the contrary, for KIRC, KIRP, LGG, PCGC, THYM, and UVM, SLC7A11 expression is positively associated with the immune score. The relationship between SLC7A11 expression and stromal/estimate score showed a similar trend. We depicted the top three cancers where the correlation coefficient between SLC7A11 expression and immune/stromal/estimate scores is largest as **Supplementary Figure S4**. Normally, tumor purity is negatively associated with immune/stromal score. Therefore, in such cancers that SLC7A11 is negatively associated with immune/stromal score, SLC7A11 may be expressed more in tumor cells instead of immune/stromal cells. Interestingly, there are major overlaps for cancers that SLC71A11 expression level is negatively associated with immune/stromal score while positively correlated with CD8$^+$ T-cell infiltration (CESC, ESCA, HNSC, LUAD, LUSC, TGCT, and THCA). In these cancers, SLC7A11 is assumed to express majorly in tumor cells and may be regulated by immune cells in stroma like CD8$^+$ T cells.

The Association Between SLC7A11 Expression and Potential Response to Immunotherapy Across Distinct Cancers

To optimize the response rate of immunotherapy, in recent years, many clinical trials tried to add chemotherapeutics in regular immunotherapy and observed a synergistic effect. Several studies have reported that chemotherapeutic drugs have the capacity to induce ferroptosis *in vivo* and *ex vivo*. Hence, it is significant to investigate whether there are correlations between SLC7A11 expression and patients' potential response to immunotherapy in the pan-cancer level. At present, the expression level of immune checkpoint markers, MSI, TMB, and neoantigen were four major indicators for

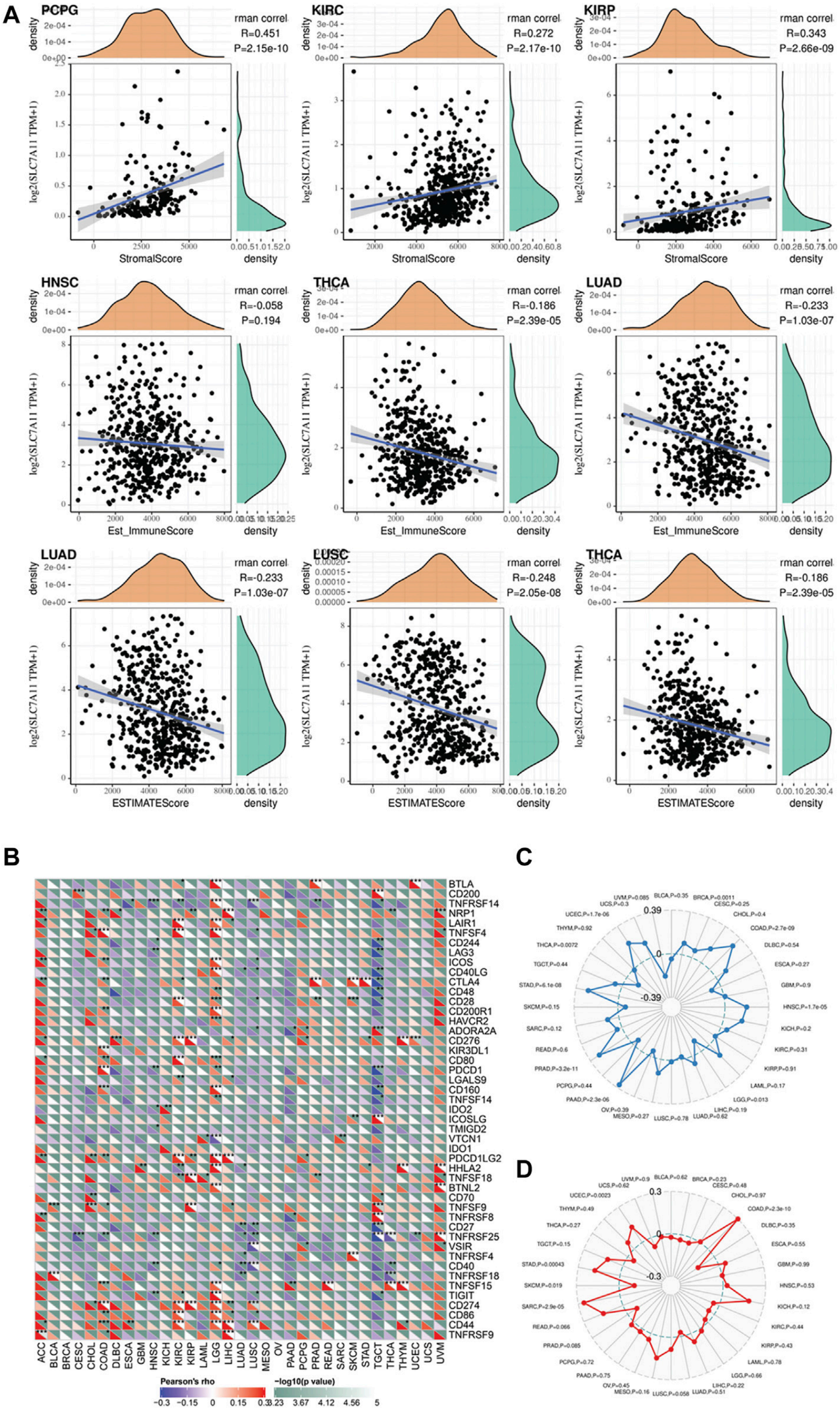


FIGURE 5 | (A) The correlation between SLC7A11 expression and immune and stromal score in distinct cancers. **(B–D)** Exploring the potential of SLC7A11 in anti-cancer immunity by evaluating its association with immune checkpoint, tumor mutation burden, and microsatellite instability across distinct cancers.

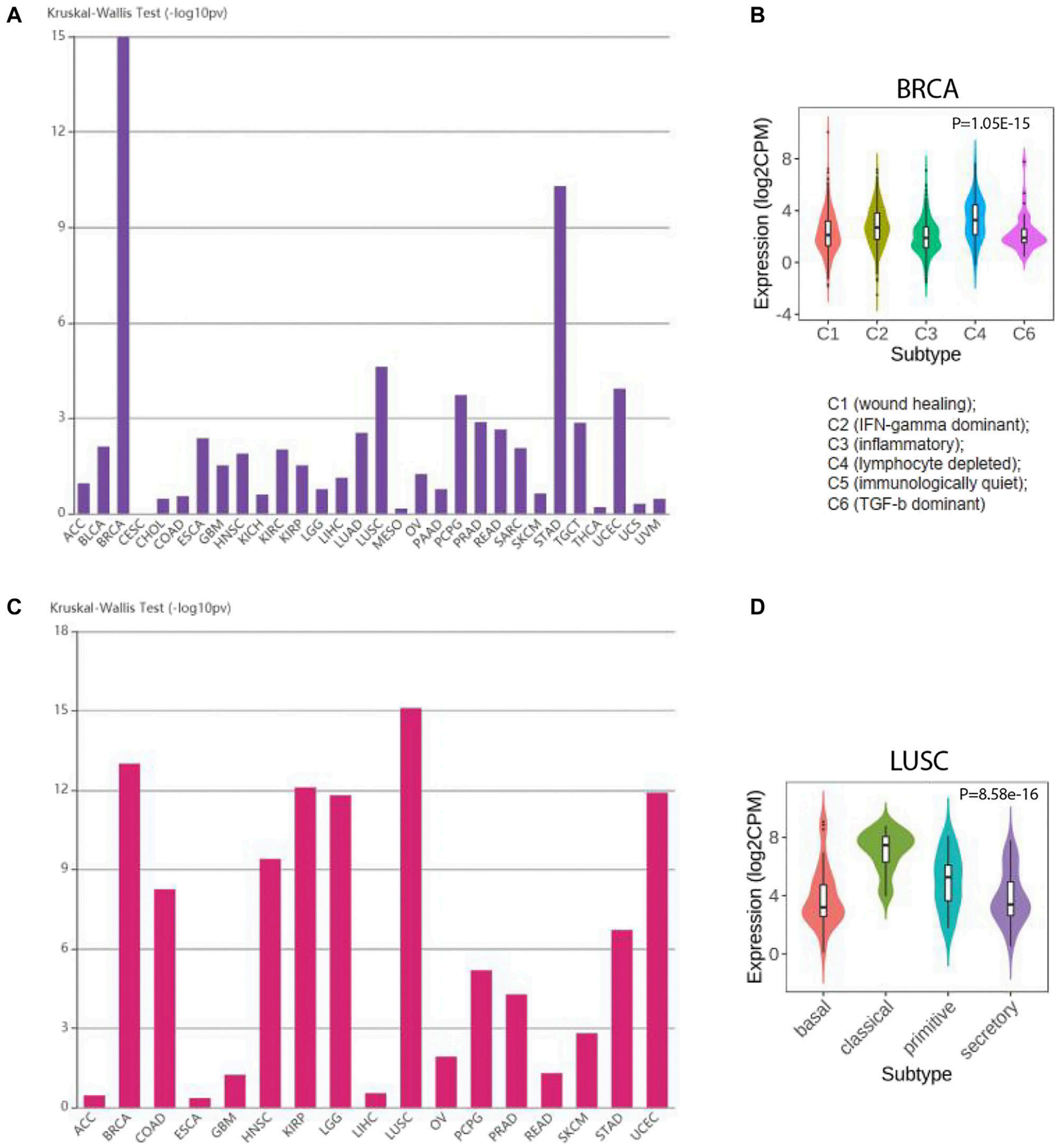


FIGURE 6 | The expression of SLC7A11 is altered in different molecular and immune subtypes of cancers. **(A, B)** SLC7A11 expression in distinct molecular subtypes for cancers. **(C, D)** SLC7A11 expression in distinct immune subtypes for cancers.

predicting patients' response to immunotherapy. Here, we analyzed the correlation between SLC7A11 expression and the level of these immunotherapeutic markers (**Figure 4B**). Two widely studied immune checkpoints, PDCD1 (PD-1) and CD274 (PD-L1), were positively correlated with SLC7A11 expression in LUAD and LGG, suggesting simultaneously targeting SLC7A11 and PD1/L1 could be beneficial in LUAD and LGG. In addition, NRP1, CD276, and VSIR showed obvious positive correlation with SLC7A11 expression in most cancers, although the drugs targeted at these checkpoints have not been applied in clinic. The TMB of STAD, PRAD, PAAD, HNSC, COAD, BRCA, and UCEC is positively correlated with SLC7A11 expression (**Figure 4C**). Among these cancer types, the SLC7A11 expression in COAD, UCEC, and STAD is also positively associated with MSI (**Figure 4D**). Besides, the expression of SLC7A11 in STAD and PRAD was positively associated with neoantigen level (**Supplementary Figure S5**).

In addition, we further studied the association between multi-omics level of SLC7A11 and immune variables (**Figures 5A–C** and **Supplementary Figure S6**). Overall, mRNA and copy numbers of SCL7A11 were correlated with most molecules involved with MHC, chemokines, and chemokine receptors; however, there was an opposite trend for the correlation between SLC7A11 methylation level and these markers, which is attributed to high methylation enrichment that caused the difficulty in transcription. Furthermore, we analyzed the correlation between SLC7A11 multi-omics expression and immune cell infiltration. The results showed that the relationship between SLC7A11 and each cell subtype was tumor specific. For example, SLC7A11 was negatively associated with activated T cells in most cancer types; however, its mRNA level was positively correlated to infiltrated activated T cells in KIRP and UVM. Accordingly, the methylation of SLC7A11 was negatively associated with activated T cells in KIRP and UVM (**Supplementary Figure S6**). Besides, the correlation between SLC7A11 and immune stimulators and inhibitors was complicated and tumor specific.

SLC7A11 Expression Altered in Distinct Molecular and Immune Subtypes

In recent years, the development of high-throughput sequencing promoted the molecular and immune subtypes for cancers. Molecular subtypes normally referred to the cancer subtypes based on clustering of differentially expressed genes. Immune subtypes were determined by the differences of immune cell infiltration in tumor microenvironment. Hence, we studied whether SLC7A11 expression was altered in different subtypes. Notably, for immune subtypes, SLC7A11 expression was significantly different in BRCA. SLC7A11 harbors the highest expression in cluster 4 (lymphocyte depleted cluster), which indicated a negative correlation for SLC7A11 expression and lymphocyte infiltration. However, for cluster 3, which was referred to inflammatory subtype, SLC7A11 harbored the lowest expression level. For molecular subtypes, SLC7A11 had the most prominent differential expression in LUSC. Specifically, classic LUSC subtype had the highest expression of SLC7A11, while secretary and basal LUSC were featured with lower SLC7A11 level (**Figure 6**).

GSEA Identified the Signal Pathways that Altered Along with Differential Expression of SLC7A11

To further explore the function of SLC7A11 in cancer biology, we performed a GSEA to investigate which signal pathways were altered along with differential expression of SLC7A11 in cancers whose prognosis was associated with SLC7A11 (**Figure 5D**). We visualized the top five gene sets enriched in SLC7A11-high or SLC7A11-low tumor samples. The results provided some valuable information. First, gene sets associated with T cellmediated immunity, T-cell receptor complex, and T-cell differentiation were enriched in HNSC samples with relatively low expression of SLC7A11, while one gene set associated with cell growth and development was enriched in HNSC samples with higher expression of SLC7A11. Second, the GSEA results in LIHC showed gene sets associated with amino acid transport and transmembrane were enriched in samples with higher SLC7A11, which may be contributed to the molecular function of SLC7A11 self. In addition, gene sets associated with G1-S phase transition and mitotic cell cycle were also enriched in LIHC samples with higher SLC7A11 expression. Third, gene sets that negatively regulate NF-kappaB transcription factor, vascular development, and angiogenesis were enriched in KICH samples with lower expression of SLC7A11. In conclusion, compared with SLC7A11-low tumor samples, cancers with higher SLC7A11 expression demonstrated decreased anti-cancer immunity, enhanced metabolic activity, and promoted cell division.

DISCUSSION

An increased number of studies have revealed an anti-cancer role of ferroptosis in recent years (Lu et al., 2017; Liang and Zhang, 2019). It is significant to investigate the implication of ferroptosis-related signature in real-world patients' tumor samples. As a core gene of ferroptosis, SLC7A11 overexpression naturally downregulates the sensitivity for ferroptosis execution (Jeong et al., 2017; Liu et al., 2019). SLC7A11 is the major target for manipulating ferroptosis *in vivo* and *in vitro* at present (Sehm et al., 2016; Xie et al., 2016; Dahlmanns et al., 2017). Hence, it is imperative to explore the expression and clinical relevance of SLC7A11 in cancer samples. This study systematically established the unfavorable role of SLC7A11 for longer survival, decreased new events, and potential response to immunotherapy in several cancers.

Thanks to the advancement of RNA-sequencing methodology, oncologists could develop a more detailed classification for cancers based on the variation in cancer transcriptome (Jeong et al., 2017; Uhlen et al., 2017). Evolutionally, tumor cell acquired unlimited proliferative capacity through the remodeling in a series of signal pathways (Nygren, 2001; Boumahdi and de Sauvage, 2020). The plasticity of transcriptome conferred tumor cells with accumulated resistance for many chemotherapeutics, which dramatically restrained the efficacy of anti-cancer treatment and compromised patients' survival (Meacham and Morrison, 2013). Mechanistically, most chemotherapeutic drugs worked out via inducing apoptosis in tumor cells (Pérez-Herrero and Fernández-Medarde, 2015). These drugs normally showed promising efficacy at the beginning, while gradually losing their ability after several rounds of administrations, which attributed to the adaptive resistance to apoptosis execution developed in tumor cells

(Goldar et al., 2015). Hence, scientists are constantly seeking a potent strategy to kill chemotherapy-resistant tumor cells. In recent years, as a novel concept, ferroptosis is defined as an iron-dependent accumulation of lipid peroxidation products to death (Hirschhorn and Stockwell, 2019). Many studies observed that chemotherapy-resistant tumor models are still sensitive to ferroptosis induction, which may become an effective weapon for the treatment of these malignancies (Liang and Zhang, 2019). Besides, ferroptosis self is also an important mechanism by which chemotherapeutic drugs kill cancer cells (Sun et al., 2016). For example, Sun et al. have demonstrated that the sensitivity to sorafenib in LIHC could be compromised by ferroptosis inhibition (Sun et al., 2016).

Unlike other non-apoptotic cell death, there are no pore-forming proteins specifically functioned in ferroptosis execution (Liang and Zhang, 2019). In this context, identification of biomarkers that reflect ferroptosis sensitivity is of great importance. As one of the most classic biomarkers of ferroptosis, SLC7A11 controls the influx of cysteine and following glutathione biosynthesis, whose overexpression is associated with ferroptostic insensitivity (Jeong et al., 2017; Koppula et al., 2018). However, few studies systematically discussed the prognostic role of SLC7A11 in cancers. In addition, one well-conducted study first proved $CD8^+$ T cells could induce the ferroptosis in cancer cells via downregulating the SLC7A11 transcription and concomitant use of immune checkpoint inhibitors could synergistically enhance the anti-cancer capacity (Wang et al., 2019). This inspired the imagination that whether intra-tumoral expression of SLC7A11 influences the immune microenvironment and potential response to immunotherapy. Hence, we elaborated a bioinformatic study in pan-cancer level to systematically explore whether the intra-tumoral expression of SLC7A11 is associated with patients' survival and its potential value in immunotherapy. Our results revealed SLC7A11 is implicated in the overall survival of several cancers, including ACC, BLCA, HNSC, KICH, KIRC, LGG, LIHC, and SKCM. Except for these cancers, SLC7A11 is associated with the DSS for PARD and KIRP. In this context, future clinical trials that aim to treat cancers by inducing ferroptosis should first consider these cancers. Our study also focused on the relationship between SLC7A11 expression and TMB, MSI, and neoantigen. Notably, STAD is the only cancer type where the SLC7A11 expression is associated with these three markers for response to immunotherapy simultaneously, which highlighted the value for combined treatment of ferroptosis induction with immunotherapy in STAD.

The present study has some strengths to declare. First, ferroptosis induction is a promising anti-cancer strategy in the near future, and our study provided many valuable and integrated suggestions for the selection of appropriate cancer types. Second, most previous studies investigated the anti-cancer function of ferroptosis in cell/animal level. Although these studies provided more detailed mechanism about the molecular pathways, our study may be more reliable because the tumor samples we selected come from human beings. Third, our study analyzed the association between SLC7A11 expression and potential response to immunotherapy in pan-cancer level, which also brought much valuable information. Certainly, this study has several limitations. On one hand, as an *in silico* research, although our study has a large sample number, appropriate external validation may still be warranted. On the other hand, while this study systematically analyzed the role of SLC7A11 in cancers' prognosis and immune signatures, the underlying mechanism still needs further investigation by future laboratory research.

In conclusion, this pan-cancer level bioinformatic study systematically elucidated the role of intra-tumoral expression of SLC7A11 in the prognoses, drug resistance, and potential immunotherapeutic response of patients with cancer. Future studies are encouraged to develop clinically approved SLC7A11-targeted drugs and organize safe and effective clinical trials.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

JH contributed to data acquisition and article drafting; HD, HL, and ZP provided technical support; QC contributed to study design and supervision. All authors read and approved the final article.

FUNDING

Scientific Research Project of Minhang District Health Committee (2019MW07).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.770857/full#supplementary-material

REFERENCES

Badgley, M. A., Kremer, D. M., Maurer, H. C., DelGiorno, K. E., Lee, H.-J., Purohit, V., et al. (2020). Cysteine Depletion Induces Pancreatic Tumor Ferroptosis in Mice. *Science* 368, 85–89. doi:10.1126/science.aaw9872

Bai, Y. T., Chang, R., Wang, H., Xiao, F. J., Ge, R. L., Wang, L. S., et al. (2018). ENPP2 Protects Cardiomyocytes from Erastin-Induced Ferroptosis. *Biochem. Biophys. Res. Commun.* 499, 44–51. doi:10.1016/j.bbrc.2018.03.113

Boumahdi, S., and de Sauvage, F. J. (2020). The Great Escape: Tumour Cell Plasticity in Resistance to Targeted Therapy. *Nat. Rev. Drug Discovdrug Discovery* 19, 39–56. doi:10.1038/s41573-019-0044-1

Burotto, M., Wilkerson, J., Stein, W. D., Bates, S. E., and Fojo, T. (2019). Adjuvant and Neoadjuvant Cancer Therapies: A Historical Review and a Rational Approach to Understand Outcomes. *Semin. Oncol.* 46, 83–99. doi:10.1053/j.seminoncol.2019.01.002

Dahlmanns, M., Yakubov, E., Chen, D., Sehm, T., Rauh, M., Savaskan, N., et al. (2017). Chemotherapeutic xCT Inhibitors Sorafenib and Erastin Unraveled with the Synaptic Optogenetic Function Analysis Tool. *Cell Death Discov.* 3, 17030. doi:10.1038/cddiscovery.2017.30

Friedmann Angeli, J. P., Krysko, D. V., and Conrad, M. (2019). Ferroptosis at the Crossroads of Cancer-Acquired Drug Resistance and Immune Evasion. *Nat. Rev. Cancer* 19, 405–414. doi:10.1038/s41568-019-0149-1

Goldar, S., Khaniani, M. S., Derakhshan, S. M., and Baradaran, B. (2015). Molecular Mechanisms of Apoptosis and Roles in Cancer Development and Treatment. *Asian Pac. J. Cancer Prev.* 16, 2129–2144. doi:10.7314/apjcp.2015.16.6.2129

Hassannia, B., Vandenabeele, P., and Vanden Berghe, T. (2019). Targeting Ferroptosis to Iron Out Cancer. *Cancer Cell* 35, 830–849. doi:10.1016/j.ccell.2019.04.002

Hause, R. J., Pritchard, C. C., Shendure, J., and Salipante, S. J. (2016). Classification and Characterization of Microsatellite Instability across 18 Cancer Types. *Nat. Med.* 22, 1342–1350. doi:10.1038/nm.4191

Hirschhorn, T., and Stockwell, B. R. (2019). The Development of the Concept of Ferroptosis. *Free Radic. Biol. Med.* 133, 130–143. doi:10.1016/j.freeradbiomed.2018.09.043

Jeong, E., Moon, S. U., Song, M., and Yoon, S. (2017). Transcriptome Modeling and Phenotypic Assays for Cancer Precision Medicine. *Arch. Pharm. Res.* 40, 906–914. doi:10.1007/s12272-017-0940-z

Koppula, P., Zhang, Y., Zhuang, L., and Gan, B. (2018). Amino Acid Transporter SLC7A11/xCT at the Crossroads of Regulating Redox Homeostasis and Nutrient Dependency of Cancer. *Cancer Commun.* 38, 12. doi:10.1186/s40880-018-0288-x

Li, Y., Chen, F., Chen, J., Chan, S., He, Y., Liu, W., et al. (2020). Disulfiram/Copper Induces Antitumor Activity against Both Nasopharyngeal Cancer Cells and Cancer-Associated Fibroblasts through ROS/MAPK and Ferroptosis Pathways. *Cancers (Basel)* 12. doi:10.3390/cancers12010138

Liang, C., and Zhang, X. (2019). Recent Progress in Ferroptosis Inducers for Cancer Therapy. *Adv. Mater.* 31, e1904197. doi:10.1002/adma.201904197

Liu, T., Jiang, L., Tavana, O., and Gu, W. (2019). The Deubiquitylase OTUB1 Mediates Ferroptosis via Stabilization of SLC7A11. *Cancer Res.* 79, 1913–1924. doi:10.1158/0008-5472.can-18-3037

Lu, B., Chen, X. B., Ying, M. D., He, Q. J., Cao, J., and Yang, B. (2017). The Role of Ferroptosis in Cancer Development and Treatment Response. *Front. Pharmacol.* 8, 992. doi:10.3389/fphar.2017.00992

Meacham, C. E., and Morrison, S. J. (2013). Tumour Heterogeneity and Cancer Cell Plasticity. *Nature* 501, 328–337. doi:10.1038/nature12624

Mou, Y., Wang, J., Wu, J., He, D., Zhang, C., Duan, C., et al. (2019). Ferroptosis, a New Form of Cell Death: Opportunities and Challenges in Cancer. *J. Hematol. Oncol.* 12, 34. doi:10.1186/s13045-019-0720-y

Nygren, P. (2001). What Is Cancer Chemotherapy? *Acta Oncologica* 40, 166–174. doi:10.1080/02841860151116204

O'Donnell, J. S., Teng, M. W. L., and Smyth, M. J. (2019). Cancer Immunoediting and Resistance to T Cell-Based Immunotherapy. *Nat. Rev. Clin. Oncol.* 16, 151–167. doi:10.1038/s41571-018-0142-8

Pérez-Herrero, E., and Fernández-Medarde, A. (2015). Advanced Targeted Therapies in Cancer: Drug Nanocarriers, the Future of Chemotherapy. *Eur. J. Pharmaceutics Biopharmaceutics* 93, 52–79. doi:10.1016/j.ejpb.2015.03.018

Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., and Hacohen, N. (2015). Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell* 160, 48–61. doi:10.1016/j.cell.2014.12.033

Sehm, T., Rauh, M., Wiendieck, K., Buchfelder, M., Eyüpoglu, I. Y., and Savaskan, N. E. (2016). Temozolomide Toxicity Operates in a xCT/SLC7a11 Dependent Manner and Is Fostered by Ferroptosis. *Oncotarget* 7, 74630–74647. doi:10.18632/oncotarget.11858

Shibata, Y., Yasui, H., Higashikawa, K., Miyamoto, N., and Kuge, Y. (2019). Erastin, a Ferroptosis-Inducing Agent, Sensitized Cancer Cells to X-ray Irradiation via Glutathione Starvation *In Vitro* and *In Vivo*. *PLoS ONE* 14, e0225931. doi:10.1371/journal.pone.0225931

Shin, C.-S., Mishra, P., Watrous, J. D., Carelli, V., D'Aurelio, M., Jain, M., et al. (2017). The Glutamate/cystine xCT Antiporter Antagonizes Glutamine Metabolism and Reduces Nutrient Flexibility. *Nat. Commun.* 8, 15074. doi:10.1038/ncomms15074

Sun, X., Niu, X., Chen, R., He, W., Chen, D., Kang, R., et al. (2016). Metallothionein-1G Facilitates Sorafenib Resistance through Inhibition of Ferroptosis. *Hepatology* 64, 488–500. doi:10.1002/hep.28574

Tang, R., Xu, J., Zhang, B., Liu, J., Liang, C., Hua, J., et al. (2020). Ferroptosis, Necroptosis, and Pyroptosis in Anticancer Immunity. *J. Hematol. Oncol.* 13, 110. doi:10.1186/s13045-020-00946-7

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., et al. (2017). A Pathology Atlas of the Human Cancer Transcriptome. *Science* 357. doi:10.1126/science.aan2507

Wang, W., Green, M., Choi, J. E., Gijón, M., Kennedy, P. D., Johnson, J. K., et al. (2019). CD8+ T Cells Regulate Tumour Ferroptosis during Cancer Immunotherapy. *Nature* 569, 270–274. doi:10.1038/s41586-019-1170-y

Wu, C., Li, M., Meng, H., Liu, Y., Niu, W., Zhou, Y., et al. (2019). Analysis of Status and Countermeasures of Cancer Incidence and Mortality in China. *Sci. China Life Sci.* 62, 640–647. doi:10.1007/s11427-018-9461-5

Xie, Y., Hou, W., Song, X., Yu, Y., Huang, J., Sun, X., et al. (2016). Ferroptosis: Process and Function. *Cell Death Differ* 23, 369–379. doi:10.1038/cdd.2015.158

Yang, G., Zheng, R.-Y., and Jin, Z.-s. (2019). Correlations between Microsatellite Instability and the Biological Behaviour of Tumours. *J. Cancer Res. Clin. Oncol.* 145, 2891–2899. doi:10.1007/s00432-019-03053-4

Yang, Y. (2015). Cancer Immunotherapy: Harnessing the Immune System to Battle Cancer. *Clin. Oncol.* 125, 3335–3337. doi:10.1172/jci83871

Zhang, H., Deng, T., Liu, R., Ning, T., Yang, H., Liu, D., et al. (2020). CAF Secreted miR-522 Suppresses Ferroptosis and Promotes Acquired Chemo-Resistance in Gastric Cancer. *Mol. Cancer* 19, 43. doi:10.1186/s12943-020-01168-8

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.

ORIGINAL RESEARCH
published: 03 January 2022
doi: 10.3389/fgene.2021.778153

Identification and Validation of Immune-Related Prognostic Genes in the Tumor Microenvironment of Colon Adenocarcinoma

*Shenghua Pan, Tingting Tang, Yanke Wu, Liang Zhang, Zekai Song and Sisi Yu**

Department of Pathology, Ruian People's Hospital, The Third Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

OPEN ACCESS

Edited by:
Luis Zapata,
Institute of Cancer Research (ICR),
United Kingdom

Reviewed by:
Xiaoxiao Sun,
University of Arizona, United States
Eleonora Lusito,
San Raffaele Telethon Institute for Gene Therapy (SR-Tiget), Italy
Ahmet Acar,
Middle East Technical University, Turkey

****Correspondence:***
Sisi Yu
zsy1737@126.com

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *16 September 2021*
Accepted: *07 December 2021*
Published: *03 January 2022*

Citation:
Pan S, Tang T, Wu Y, Zhang L, Song Z and Yu S (2022) Identification and Validation of Immune-Related Prognostic Genes in the Tumor Microenvironment of Colon Adenocarcinoma. Front. Genet. 12:778153. doi: 10.3389/fgene.2021.778153

The tumor microenvironment (TME) has been shown to be involved in angiogenesis, tumor metastasis, and immune response, thereby affecting the treatment and prognosis of patients. This study aims to identify genes that are dysregulated in the TME of patients with colon adenocarcinoma (COAD) and to evaluate their prognostic value based on RNA omics data. We obtained 512 COAD samples from the Cancer Genome Atlas (TCGA) database and 579 COAD patients from the independent dataset (GSE39582) in the Gene Expression Omnibus (GEO) database. The immune/stromal/ESTIMATE score of each patient based on their gene expression was calculated using the ESTIMATE algorithm. Kaplan–Meier survival analysis, Cox regression analysis, gene functional enrichment analysis, and protein–protein interaction (PPI) network analysis were performed. We found that immune and stromal scores were significantly correlated with COAD patients' overall survival (log rank $p < 0.05$). By comparing the high immune/stromal score group with the low score group, we identified 688 intersection differentially expressed genes (DEGs) from the TCGA dataset (663 upregulated and 25 downregulated). The functional enrichment analysis of intersection DEGs showed that they were mainly enriched in the immune process, cell migration, cell motility, Toll-like receptor signaling pathway, and PI3K-Akt signaling pathway. The hub genes were revealed by PPI network analysis. Through Kaplan–Meier and Cox analysis, four TME-related genes that were significantly related to the prognosis of COAD patients were verified in GSE39582. In addition, we uncovered the relationship between the four prognostic genes and immune cells in COAD. In conclusion, based on the RNA expression profiles of 1091 COAD patients, we screened four genes that can predict prognosis from the TME, which may serve as candidate prognostic biomarkers for COAD.

Keywords: tumor microenvironment, immune/stromal scores, colon adenocarcinoma, protein–protein interaction network, OMICS data

---

Abbreviations: COAD, colon adenocarcinoma; KM, Kaplan–Meier; TME, tumor microenvironment; TCGA, the Cancer Genome Atlas; PPI, protein–protein interaction network.

INTRODUCTION

Colorectal cancer (CRC) is a common malignant gastrointestinal tumor worldwide (Siegel et al., 2017; Siegel et al., 2020). Colon adenocarcinoma (COAD) is the most common histological type of CRC (Barresi et al., 2015). According to GLOBOCAN 2018, CRC is the malignant tumor with the third highest incidence and the second highest mortality. It is estimated that there were more than 1.8 million new cases and 881,000 deaths caused by CRC in 2018 (Bray et al., 2018). In addition, the incidence of CRC among young adults is increasing (Benson et al., 2017), which brings a huge health burden to human beings worldwide. The prognosis of CRC varies in different countries around the world. The 5-year relative survival rate of CRC in high-income countries is close to 65%, while in low-income countries, it is less than 50% (Brenner et al., 2014). Despite the continuous development of treatment methods such as operation, chemotherapy agents, and radiotherapy, the prognosis of CRC has not been significantly improved. Recently, immunotherapy has become a promising therapeutic method for CRC patients. Unfortunately, current clinical trials show that only a few people can benefit from immunotherapy; thus, finding biomarkers that can indicate treatment response and prognosis has become an urgent problem (Piawah and Venook, 2019).

The tumor microenvironment (TME) has been proven to be involved in angiogenesis, tumor metastasis, and immune response, thereby affecting the treatment and prognosis of patients (Qi and Wu, 2019). The TME is composed of immune cells (T cells, macrophages, etc.), stromal cells (endothelial cells, etc.), and extracellular components (cytokines, hormones, etc.). Immune and stromal cells are reported to be the key carriers for the tumor microenvironment to perform multiple biological functions. For CRC, researchers have confirmed the prognostic role of tumor-infiltrating immune cells in the TME (Galon et al., 2006). Immune and stromal classification of CRC has been found to be associated with molecular subtypes and precision immunotherapy (Becht et al., 2016; Micke et al., 2021). Therefore, understanding the immune status of the TME is greatly significant for improving the treatment and prognosis of COAD.

Based on the gene expression value in the TME, Yoshihara et al. (2013) constructed a new algorithm and called it "ESTIMATE" to evaluate the proportion of stromal and immune cells in tumor tissues. Through this method, researchers have discovered diagnostic or prognostic markers of glioblastoma, cervical squamous cell carcinoma, bladder cancer, gastric cancer, etc. (Jia et al., 2018; Luo et al., 2019; Pan et al., 2019; Wang et al., 2019), as well as tumor immune-related therapeutic targets. However, the prognostic value of the TME of COAD has not yet been elucidated. In this study, we used the COAD gene expression profile data from TCGA to calculate the immune/stromal scores of COAD patients using ESTIMATE and explored the correlation of these scores with the clinical characters and overall survival of COAD patients. Subsequently, we investigated the potential prognostic genes in the TME of COAD.

TABLE 1 | Summary of patient demographics and clinical characteristics.

Characteristic	TCGA	GSE39582
Age (years)	69 (31–90)	69 (22–97)
Gender		
Female	244	260
Male	266	319
Unknown	2	
Vital status		
Living	379	385
Dead	108	194
Unknown	25	
M stage		
M0	370	496
M1	72	61
Unknown	70	22
N stage		
N0	305	311
N1	114	136
N2	91	100
Unknown	2	32
T stage		
T1	11	12
T2	85	48
T3	349	376
T4	64	119
Unknown	3	24
Tumor stage		
Stage I	82	37
Stage II	205	269
Stage III	139	209
Stage IV	72	60
Unknown	24	4

MATERIALS AND METHODS

Patients and Gene Expression Data

We collected the RNA expression data of 512 COAD patients from the TCGA database (https://cancergenome.nih.gov/). Clinical and pathological characteristics, including gender, age, and pathological tumor staging of all 512 COAD patients, are listed in **Table 1**. The log2 transformed FPKM values were used for gene expression analysis. Due to the lack of survival information of 25 patients, we used the expression and survival data of 487 patients for further analysis. In order to verify the prognostic value of genes in COAD and the relationship between genes and immune cell behavior, we obtained another group of 587 COAD patients with RNA expression profiles and clinical characters (GSE39582) from the Gene Expression Omnibus (GEO) database.

Calculating Immune/Stromal Scores and Survival Analysis

Based on the ESTIMATE algorithm in the R program (3.5.3), we obtained the immune/stromal/ESTIMATE score of each sample (Yoshihara et al., 2013). Subsequently, the degree of infiltration of immune cells was quantified by Single Sample Gene Set Enrichment Analysis (ssGSEA) (Hänzelmann et al., 2013; Xiao et al., 2020). In order to determine the optimal cut-off value of the

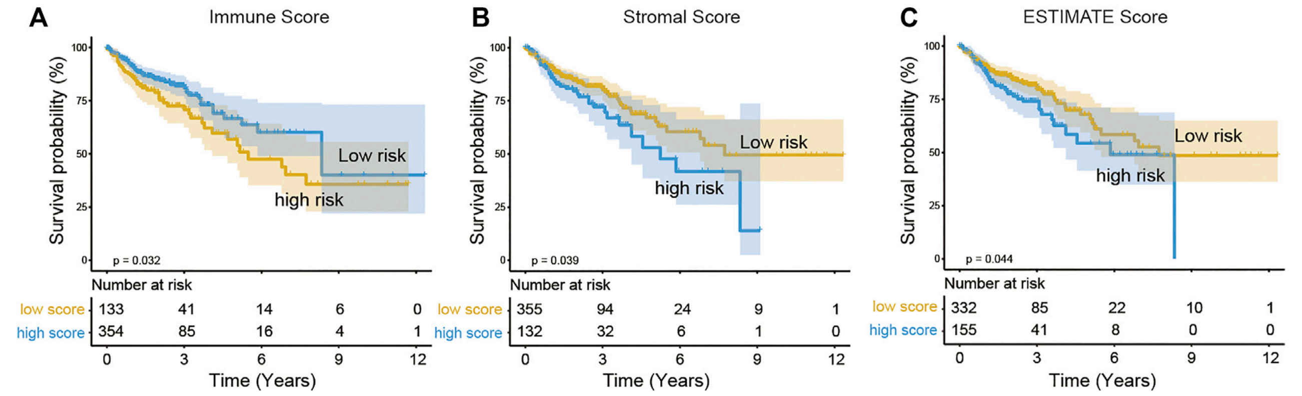


FIGURE 1 | Association of immune scores **(A),** stromal scores **(B),** and ESTIMATE scores **(C)** with COAD overall survival. The COAD cases were divided into two groups based on their immune scores or stromal scores or ESTIMATE scores. Kaplan–Meier survival curve of OS between high and low immune/stromal/ESTIMATE score groups.

immune/stromal score to classify participants into high-/low-score groups, R packages including "maxstat" and "survival" were used (Hothorn and Zeileis, 2008). Kaplan–Meier (KM) analysis was performed to explore the prognostic performance of the immune/stromal/ESTIMATE score, and the log rank p value was computed and showed on the survival curves. To understand the correlation of the tumor stage with the immune/stromal score, one-way ANOVA was used to test differences. The differentially expressed genes (DEGs) with a $|\text{fold change}| > 1.5$ and a p value <0.05 were found by SAM test, which was a statistical technique based on a t-test in R software (Tusher et al., 2001). KM and COX regression analysis were used to further evaluate the relationship between the DEGs and over survival of COAD in the TCGA and GSE39582 datasets.

Functional Prediction and PPI Network Analysis

ClueGo of the Cytosccape plug-in (Bindea et al., 2009) was performed to predict the biological function of DEGs, which could cluster genes using Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG). The protein–protein interaction (PPI) network was constructed through the STRING database (von Mering et al., 2005), and these selected genes required a confidence score ≥ 0.4 and a maximum number of interactors = 0. The visualization analysis of PPI was completed using Cytoscape software (Shannon et al., 2003). The Network Analyzer plug-in of Cytoscape was used to analyze the degree distribution of genes.

RESULTS

Prognostic Correlation Analysis of Immune/Stromal Scores in COAD

As shown in **Table 1**, the median age of the 512 COAD patients in TCGA was 69, males outnumbered females, and patients without lymph node metastasis and distant metastasis (stage I and II) accounted for the majority. A total of 487 patients with complete survival information and gene expression data were studied. From the gene expression profiles, we identified 17,590 expressed genes in the 487 COAD samples. We acquired the immune, stromal, and ESTIMATE scores of each COAD patient using ESTIMATE (**Supplementary Table S1**). The immune scores of the 487 COAD patients ranged from −1,262.3 to 2,598.7, the stromal scores ranged from −2,543.4 to 1,622.9, and the ESTIMATE scores ranged from −3,579.2 to 3,689.2. To investigate the potential correlation between the prognosis of COAD and the immune/stromal/ESTIMATE score, we divided patients into low-score or high-score groups by the cut-off value selected by maximally selected rank statistics in the R maxstat package. Kaplan–Meier analysis revealed that the high immune score group with a score higher than −202.9 was significantly correlated with a better prognosis than the low immune score group (median survival 8.33 vs. 5.49 years, log rank $p = 0.03$, **Figure 1A**). Based on −382.6/49.5 as the selected cut-off value, the high-score groups of the stromal/ESTIMATE scores had a shorter survival (stromal: median survival 5.23 vs. 7.73 years; ESTIMATE: median survival 5.85 vs. 7.73 years, log rank $p < 0.05$, **Figures 1B,C**).

Subsequently, we investigated the association of immune/stromal/ESTIMATE scores with the COAD tumor stage and pathologic T, N, and M stages by one-way ANOVA test. As shown in **Figure 2** and **Supplementary Figure S1**, the immune scores were significantly associated with the tumor stage and pathologic M stage ($p < 0.05$, **Figure 2A**, **Supplementary Figure S1G**). But the stromal scores or ESTIMATE scores were not correlated with that of COAD ($p > 0.05$, **Figures 2B,C**, **Supplementary Figure S1**). Then we ran Tukey's HSD test to compare the scores between different tumor stages (**Figures 2D–F**) and found out that the group means of immune scores owned a significantly different value between tumor stages IV and II ($p < 0.05$, **Figure 2D**).

Differentially Expressed Gene Analysis in the TME of COAD

After obtaining the immune scores, we performed differentially expressed gene analysis based on the high- ($n = 354$) and low-

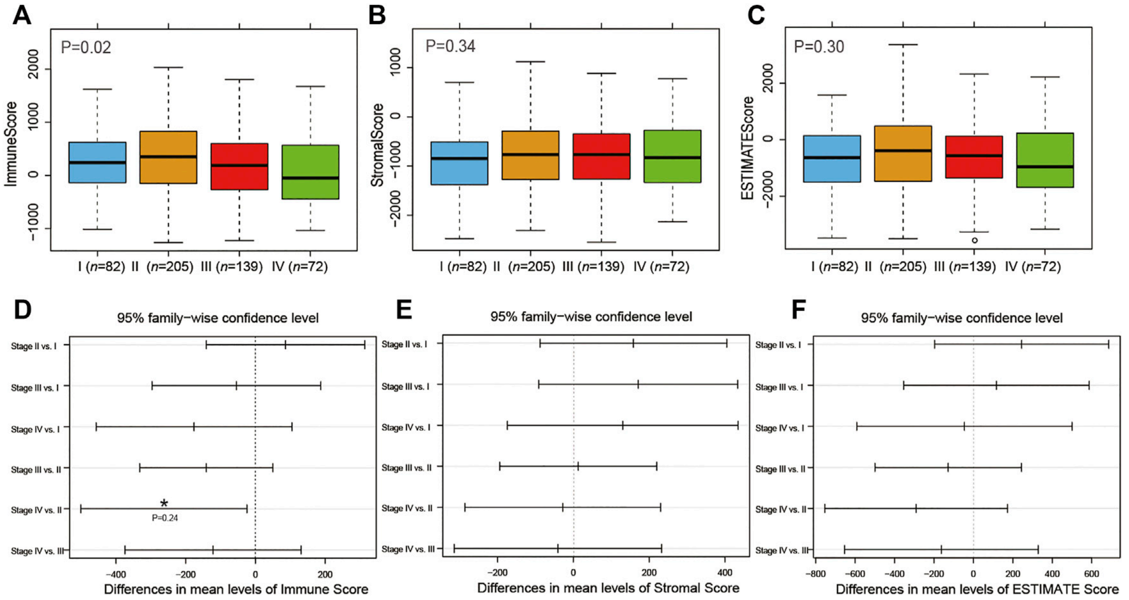


FIGURE 2 | Association of tumor stage with immune **(A)**, stromal **(B)**, and ESTIMATE **(C)** scores. Tukey's HSD test to compare the differences between different tumor stages with immune **(D)**, stromal **(E)**, and ESTIMATE **(F)** scores.

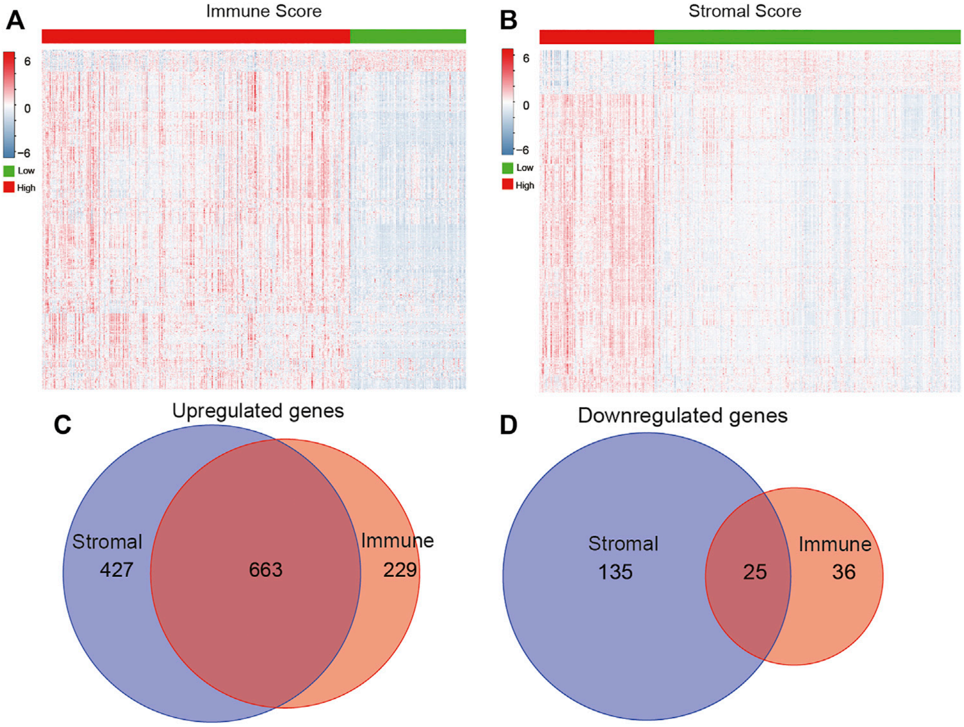


FIGURE 3 | Identification of DEGs based on immune/stromal scores. Heatmap of DEGs from the low vs. high immune **(A)**/stromal **(B)** score groups (|fold change| >1.5, $p < 0.05$). Venn diagrams showed the number of overlapped up- or downregulated DEGs in immune score **(C)** and stromal score **(D)** groups.

score (n = 133) groups. A total of 953 DEGs were identified, of which 892 DEGs were upregulated genes and 61 DEGs were downregulated genes (**Figure 3A**). Meanwhile, there were 1,090 upregulated and 160 downregulated DEGs according to the comparison between high stromal score (n = 132) and low stromal score (n = 355) groups (**Figure 3B**). Venn diagrams

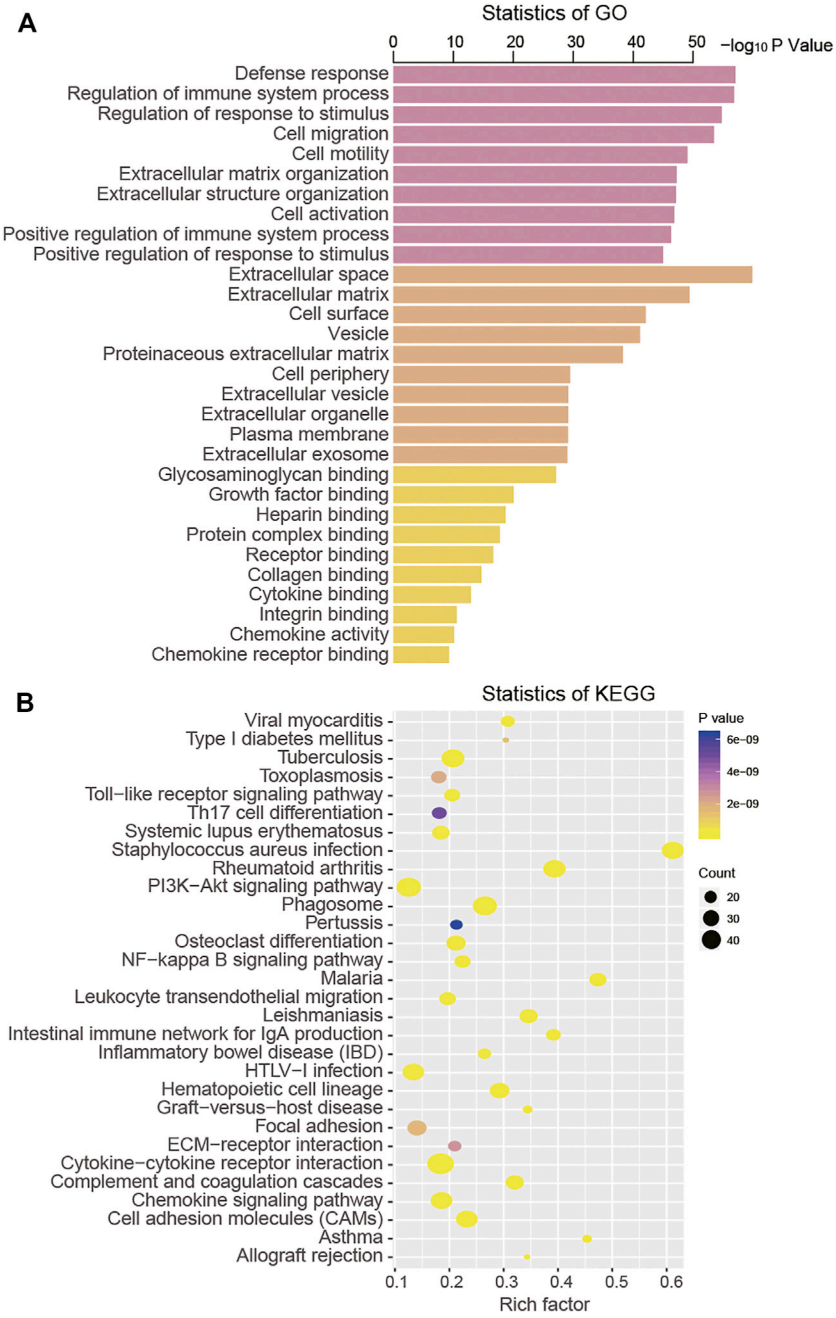


FIGURE 4 | Functional analysis of intersection DEGs by GO **(A)** and KEGG **(B)**.

indicated 663 overlapping upregulated genes and 25 overlapping downregulated genes in both immune and stromal groups (**Figures 3C,D**). Further analysis focused on the common DEGs.

Through the ClueGO annotation in Cytoscape software, we conducted GO and KEGG analysis to predict the function of the 688 intersection DEGs and found that these genes were mainly clustered in 922 GO terms and 44 KEGG pathways (**Supplementary Table S2**). From the aspects of biological processes (BPs), we found that these intersection DEGs were mainly enriched in cell migration, cell motility, and regulation of the immune system process. From the aspects of the cellular component (CC), these DEGs were primarily clustered in the extracellular space and the extracellular matrix. At the level of molecular function (MF), they were mainly associated with glycosaminoglycan binding, growth factor binding, and heparin binding (**Figure 4A**). The KEGG pathway analysis result suggested that these DEGs were mainly enriched in the

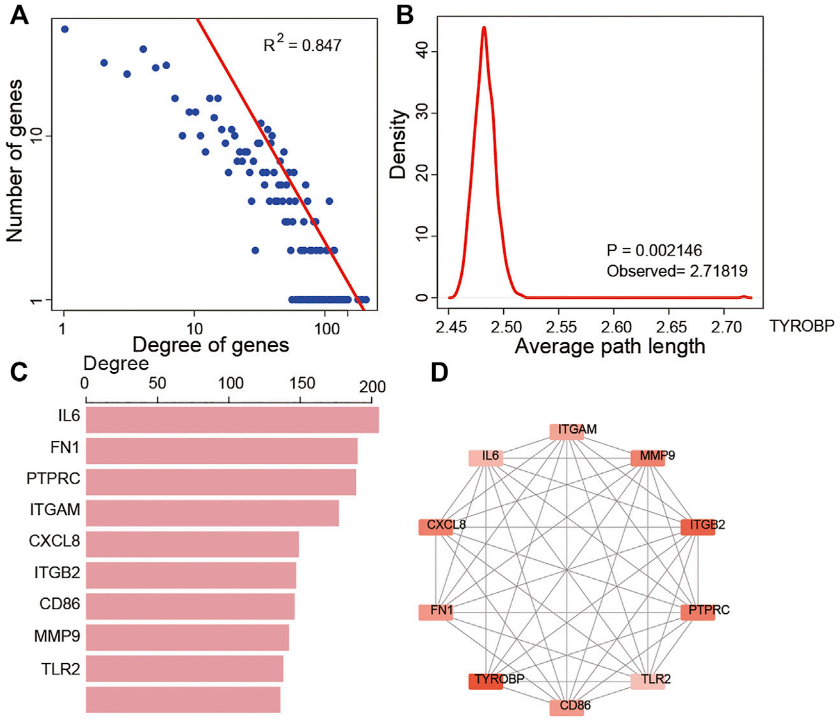


FIGURE 5 | Topological features of DEGs in the PPI network. **(A)** View of the PPI network. **(A)** Degree distributions of the PPI network. **(B)** All degrees followed a power-law distribution and average path length distributions of the real network and 1,000 times random networks. **(C)** Key DEGs in the PPI network with top 10 degree distributions. **(D)** Subnetwork of the top 10 genes.

toll-like receptor signaling pathway, cell adhesion molecules (CAMs), and the PI3K-Akt signaling pathway (**Figure 4B**).

PPI Network Construction of DEGs

The PPI network of these 688 DEGs was constructed based on the STRING database, and a total of 664 nodes and 10,015 interactions were detected (**Supplementary Figure S2**). We further analyzed the node degree in PPI and found that they obeyed the power-law distribution, indicating that the network was scale-free, similar to most biological networks (**Figure 5A**). In addition, we calculated the average path length of the PPI network, which showed that the characteristic path length of the network was much longer than that of the random network (1,000 times that of the random network, $p = 0.002$, **Figure 5B**). The most highly connected intersection DEGs were identified. Among these, IL6, FN1, PTPRC, ITGAM, CXCL8, ITGB2, CD86, MMP9, TLR2, and TYROBP were the top ten with 205, 190, 189, 177, 149, 147, 146, 142, 138, and 136 nodes, respectively (**Figure 5C**). So we grabbed the subnetwork of the 10 genes and found that most of them were interactive and are highly expressed in the high immune score group (**Figure 5D**).

Prognostic Value of DEGs in COAD

The association of intersection DEGs with OS of COAD was evaluated by Kaplan–Meier and Cox regression analysis in the 487 COAD cases. Among 668 intersection DEGs, a total of 38 genes were associated with the OS ($p < 0.05$, **Supplementary Table S3**). Among the 38 prognostic DEGs, patients with high expression of CPA3, MMP12, MMP1, CXCL8, and TSPAN11 were associated with longer OS than those with low expression, while the upregulated expression of other genes were associated with unfavorable outcomes.

To further validate above results, an independent dataset including 579 patients from the GEO database (accession number GSE39582) were used to verify the prognostic genes of COAD. As a result, we found that four genes out of a total of 38 identified genes from the TCGA were significantly associated with COAD survival. These four prognostic genes were VIM, SIGLEC1, ARL4C, and CPA3. From **Figure 6**, we observed that high expression of VIM, SIGLEC1, and ARL4C and low expression of CPA3 were associated with poor prognosis.

Relationships Between the Four Prognostic Genes and Immune Cell Behavior in COAD

Next, to uncover the potential role of genes in tumor immunology, ssGSEA analysis unearthed the ratio of immune cells, and then we analyzed the correlation between the expression of these genes and immune cells by Pearson test and presented it with heatmaps based on TCGA (**Figure 7A**) and GSE39582 (**Figure 7B**) datasets. Interestingly, we observed that CPA3 was highly correlated only with mast cells (**Figures**

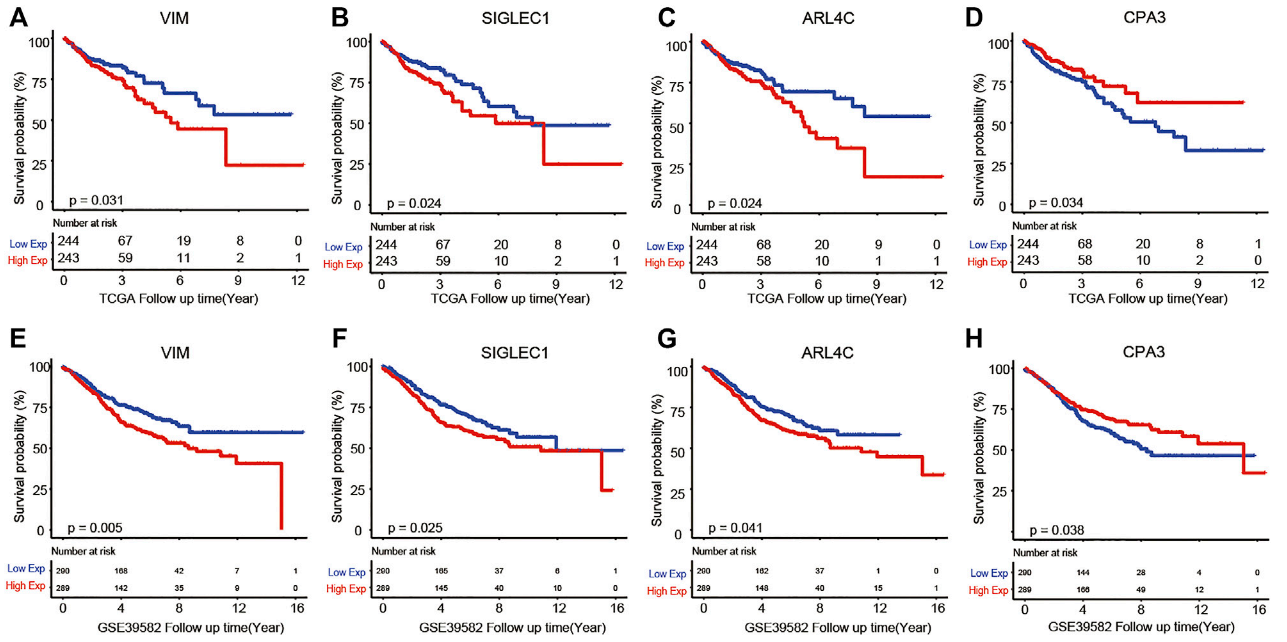


FIGURE 6 | Kaplan–Meier analysis results of VIM **(A,E)**, SIGLEC1 **(B,F)**, ARL4C **(C,G)**, and CPA3 **(D,H)** in TCGA and GSE39582 datasets.

7C,D), while VIM, SIGLEC1, and ARL4C were highly correlated with macrophage, natural killer cell, regulatory T cell, T follicular helper cell, and Type 1 T helper cell in TCGA (**Supplementary Figure S3**) and GSE39582 (**Supplementary Figure S4**), simultaneously (Pearson coefficient>0.6, $p < 0.001$). These results suggested that the four prognostic genes may participate in cancer progression by regulating the level of immune cells in COAD (Pearson coefficient>0.6, $p < 0.001$).

DISCUSSION

COAD is a heterogeneous malignant tumor with widely variable prognosis (Barresi et al., 2015). Therefore, new prognostic biomarkers and theraputic methods are needed. Recently, immunotherapy has brought great hope to COAD patients, but its limited effectiveness and drug resistance are still major challenges. The TME has been reported to be implicated in the development of various tumors (Balkwill et al., 2012; Hui and Chen, 2015) and affect the treatment and prognosis of patients, but research on the TME of COAD is rare. In this study, we identified the correlation of the immune/stromal scores with the survival of COAD. Moreover, we screened out a total of 688 DEGs from high vs. low immune/stromal score groups and found four genes with prognostic value in the TME, which have the potential ability to serve as molecular biomarkers of COAD.

After analyzing the profiles of 487 COAD patients through the ESTIMATE algorithm, we found that a high immune score was correlated with a high overall survival rate of COAD, which was identical to the results reported in other tumors such as hepatocellular carcinoma (Pan et al., 2020), prostate cancer (Sun et al., 2020), and endometrial cancer (Chen et al., 2020). This correlation result indicates that immune cell infiltration is beneficial to the prognosis of COAD. Similarly, Michael J et al. found that high tumor-associated macrophage infiltration in CRC was associated with better prognosis (Cavnar et al., 2017). Franck et al. discovered that cytotoxic (CD8) and memory (CD45RO) T cells could predict better clinical outcomes of CRC patients (Pagès et al., 2009). Other researchers have confirmed that immune cells are prognostic factors for CRC (Galon et al., 2006). On the other hand, our study also discovered that the high stromal/ESTIMATE scores were correlated with poor prognosis, indicating that the stromal cells in the TME are indicators of unfavorable clinical outcome for CRC. Consistent with our findings, the high expression of cancer-associated fibroblasts (CAFs, one group of stromal cells) was found to be associated with the poor prognosis of untreated CRC (Isella et al., 2015). Moreover, some researchers have found that stromal cells in the TME of colon cancer have a key role in inhibiting tumor immune response and enhancing tumor malignant progression (O'Malley et al., 2018).

When analyzing the correlation between clinical parameters and immune/stromal scores, we found significant differences in the immune scores of COAD patients with different tumor stages. Then, we obtained 688 DEGs through analyzing DEGs that appear in both groups with high and low immune/stromal scores. Function analysis found that these intersection DEGs were enriched in the immune system process, cell migration, cell motility, growth factor binding and extracellular matrix, Toll-like receptor signaling pathway, NF-kappa B signaling pathway, and PI3K-Akt signaling pathway, which were closely related to tumor metastasis. Moreover, we identified that the most highly connected intersection DEGs in the TME of COAD were IL6. IL6 (Interleukin-6), one of the major cytokines in the TME, has

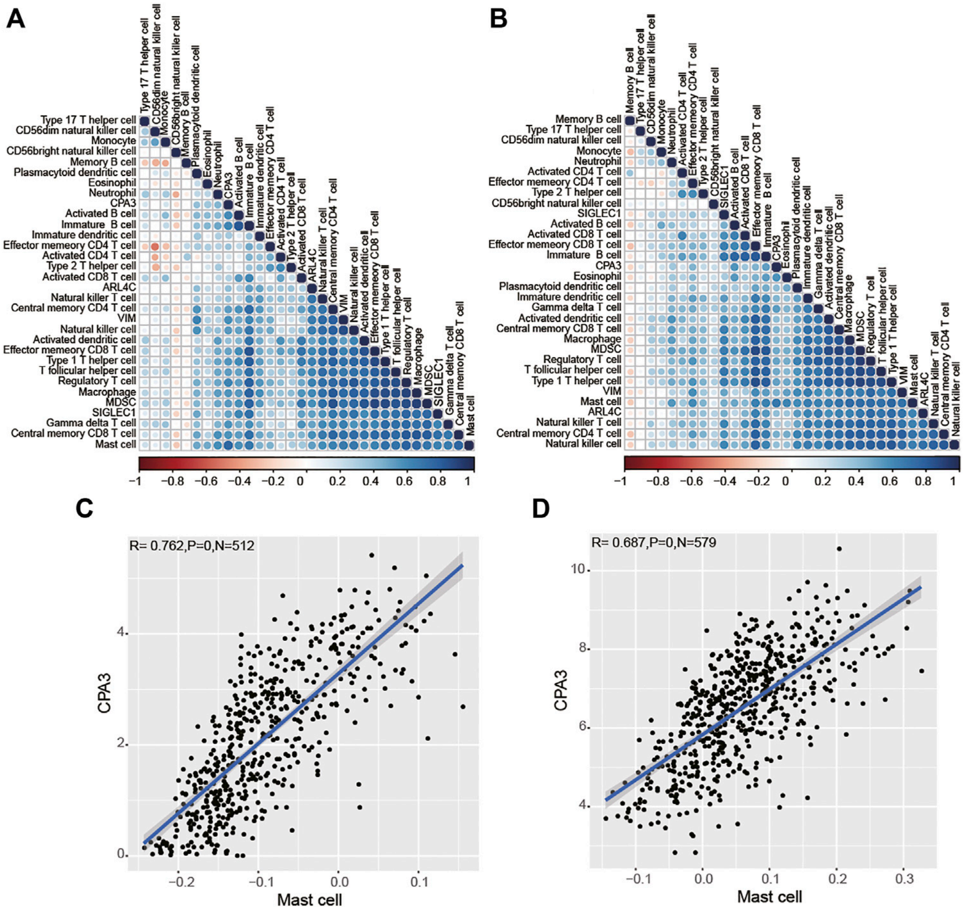


FIGURE 7 | Relationships between the four prognostic genes and immune cells based on TCGA **(A)** and GSE39582 **(B)** dataset analysis. CPA3 was highly correlated with Mast cells in TCGA **(C)** and GSE39582 **(D)** datasets.

been reported to promote tumor progression including apoptosis, proliferation, invasiveness, and metastasis *via* regulating multiple key cell signaling pathways (Kumari et al., 2016). All the above findings indicated that the TME of COAD had an important role in tumor progress and outcome.

Subsequently, Kaplan–Meier and COX analysis found that 38 TME-related DEGs were significantly correlated with the OS of COAD patients from TCGA database and validated four genes (VIM, SIGLEC1, ARL4C, and CPA3) in the GEO dataset. VIM (vimentin) gene encodes type III intermediate filament protein and is expressed in most cell types, particularly tumor cells. VIM has an important function of regulating cell migration (Battaglia et al., 2018). It has been reported that the abnormally high expression of vimentin in various epithelial cancers including prostate cancer, gastrointestinal tumors, and breast cancer is closely related to tumor growth, invasion, and poor prognosis (Satelli and Li, 2011). In CRC cells, researchers found that siRNA knockdown of VIM expression could reduce cell migration and invasiveness (McInroy and Määttä, 2007). Consistent with the results of the above studies, our study found that high expression of VIM indicated poor prognosis for patients with COAD. SIGLEC1, also known as CD169, encodes a type I transmembrane protein expressed on macrophages. Studies have shown that CD169$^+$ macrophages are a favorable prognostic indicator for bladder cancer (Asano et al., 2018) and hepatocellular carcinoma (Zhang et al., 2016). These results in other tumors are contrary to this article, so it is necessary to further clarify the prognostic significance and specific mechanisms of SIGLEC1. ARL4C (ADP-ribosylation factor-like protein 4C) is a target gene for both Wnt/β-catenin and epidermal growth factor/Ras signaling, and its expression is reported to promote cellular migration and proliferation, thereby indicating its involvement in tumorigenesis. It has been found that ARL4C is overexpressed in colorectal cancers and plays a pivotal role in the progression of CRC (Fujii et al., 2015; Chen et al., 2016). CPA3 (carboxypeptidase A3) is a member of the metallocarboxypeptidase family and can be expressed in many cell types, especially basophils and mast cells. There are few studies on the expression of CPA3 in tumors and its prognostic significance. Our study found that CPA3 was a protective factor of COAD, and high expression of CPA3 was associated with better survival rates.

In conclusion, this study provides an attempt at understanding the role of immune/stromal cells and genes in the TME of COAD and confirms that the composition of TME affects the clinical outcomes of COAD patients. Moreover, four TME-related genes have been identified, which could be used as new prognostic biomarkers and targets for immunotherapy.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. The data can be found here. Publicly available datasets can be found here: https://gdc-hub.s3.us-east-1.amazonaws.com/download/TCGA-COAD.htseq_fpkm-uq.tsv.gz; https://ftp.ncbi.nlm.nih.gov/geo/series/GSE39nnn/GSE39582/matrix/GSE39582_series_matrix.txt.gz.

AUTHOR CONTRIBUTIONS

The authors contributed in the following way: SP: data collection, data analysis, interpretation, and drafting; SY: study design, study supervision, and final approval of the manuscript; TT, YW, LZ, and ZS: technical support and critical revision of the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the Wenzhou Basic Scientific Research Project (grant number: Y2020956).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.778153/full#supplementary-material

REFERENCES

Asano, T., Ohnishi, K., Shiota, T., Motoshima, T., Sugiyama, Y., Yatsuda, J., et al. (2018). CD 169-positive Sinus Macrophages in the Lymph Nodes Determine Bladder Cancer Prognosis. *Cancer Sci.* 109, 1723–1730. doi:10.1111/cas.13565

Balkwill, F. R., Capasso, M., and Hagemann, T. (2012). The Tumor Microenvironment at a Glance. *J. Cel. Sci.* 125, 5591–5596. doi:10.1242/jcs.116392

Barresi, V., Reggiani Bonetti, L., Ieni, A., Caruso, R. A., and Tuccari, G. (2015). Histological Grading in Colorectal Cancer: New Insights and Perspectives. *Histol. Histopathol* 30, 1059–1067. doi:10.14670/HH-11-633

Battaglia, R. A., Delic, S., Herrmann, H., and Snider, N. T. (2018). Vimentin on the Move: New Developments in Cell Migration. *F1000Res* 7, F1000Res. doi:10.12688/f1000research.15967.1

Becht, E., de Reyniès, A., Giraldo, N. A., Pilati, C., Buttard, B., Lacroix, L., et al. (2016). Immune and Stromal Classification of Colorectal Cancer Is Associated with Molecular Subtypes and Relevant for Precision Immunotherapy. *Clin. Cancer Res.* 22, 4057–4066. doi:10.1158/1078-0432.ccr-15-2879

Benson, A. B., 3rd, Venook, A. P., Cederquist, L., Chan, E., Chen, Y.-J., Cooper, H. S., et al. (2017). Colon Cancer, Version 1.2017, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Canc Netw.* 15, 370–398. doi:10.6004/jnccn.2017.0036

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape Plug-In to Decipher Functionally Grouped Gene Ontology and Pathway Annotation Networks. *Bioinformatics* 25, 1091–1093. doi:10.1093/bioinformatics/btp101

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a Cancer J. clinicians* 68, 394–424. doi:10.3322/caac.21492

Brenner, H., Kloor, M., and Pox, C. P. (2014). Colorectal Cancer. *The Lancet* 383, 1490–1502. doi:10.1016/s0140-6736(13)61649-9

Cavnar, M. J., Turcotte, S., Katz, S. C., Kuk, D., Gönen, M., Shia, J., et al. (2017). Tumor-Associated Macrophage Infiltration in Colorectal Cancer Liver Metastases Is Associated with Better Outcome. *Ann. Surg. Oncol.* 24, 1835–1842. doi:10.1245/s10434-017-5812-8

Chen, P., Yang, Y., Zhang, Y., Jiang, S., Li, X., and Wan, J. (2020). Identification of Prognostic Immune-Related Genes in the Tumor Microenvironment of Endometrial Cancer. *Aging* 12, 3371–3387. doi:10.18632/aging.102817

Chen, X., Su, Z., Wang, S., and Xu, H. (2016). Clinical and Prognostic Significance of Arl4c Expression in Colorectal Cancer. *Cbm* 16, 253–257. doi:10.3233/cbm-150562

Fujii, S., Matsumoto, S., Nojima, S., Morii, E., and Kikuchi, A. (2015). Arl4c Expression in Colorectal and Lung Cancers Promotes Tumorigenesis and May Represent a Novel Therapeutic Target. *Oncogene* 34, 4834–4844. doi:10.1038/onc.2014.402

Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., et al. (2006). Type, Density, and Location of Immune Cells within Human Colorectal Tumors Predict Clinical Outcome. *Science* 313, 1960–1964. doi:10.1126/science.1129139

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7

Hothorn, T., and Zeileis, A. (2008). Generalized Maximally Selected Statistics. *Biometrics* 64, 1263–1269. doi:10.1111/j.1541-0420.2008.00995.x

Hui, L., and Chen, Y. (2015). Tumor Microenvironment: Sanctuary of the Devil. *Cancer Lett.* 368, 7–13. doi:10.1016/j.canlet.2015.07.039

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, 312–319. doi:10.1038/ng.3224

Jia, D., Li, S., Li, D., Xue, H., Yang, D., and Liu, Y. (2018). Mining TCGA Database for Genes of Prognostic Value in Glioblastoma Microenvironment. *Aging* 10, 592–605. doi:10.18632/aging.101415

Kumari, N., Dwarakanath, B. S., Das, A., and Bhatt, A. N. (2016). Role of Interleukin-6 in Cancer Progression and Therapeutic Resistance. *Tumor Biol.* 37, 11553–11572. doi:10.1007/s13277-016-5098-7

Luo, Y., Zeng, G., and Wu, S. (2019). Identification of Microenvironment-Related Prognostic Genes in Bladder Cancer Based on Gene Expression Profile. *Front. Genet.* 10, 1187. doi:10.3389/fgene.2019.01187

McInroy, L., and Määttä, A. (2007). Down-regulation of Vimentin Expression Inhibits Carcinoma Cell Migration and Adhesion. *Biochem. Biophysical Res. Commun.* 360, 109–114. doi:10.1016/j.bbrc.2007.06.036

Micke, P., Strell, C., Mattsson, J., Martín-Bernabé, A., Brunnström, H., Huvila, J., et al. (2021). The Prognostic Impact of the Tumour Stroma Fraction: A Machine Learning-Based Analysis in 16 Human Solid Tumour Types. *EBioMedicine* 65, 103269. doi:10.1016/j.ebiom.2021.103269

O'Malley, G., Treacy, O., Lynch, K., Naicker, S. D., Leonard, N. A., Lohan, P., et al. (2018). Stromal Cell PD-L1 Inhibits CD8$^+$ T-Cell Antitumor Immune Responses and Promotes Colon Cancer. *Cancer Immunol. Res.* 6, 1426–1441. doi:10.1158/2326-6066.cir-17-0443

Pagès, F., Kirilovsky, A., Mlecnik, B., Asslaber, M., Tosolini, M., Bindea, G., et al. (2009). *In Situ* cytotoxic and Memory T Cells Predict Outcome in Patients with Early-Stage Colorectal Cancer. *Jco* 27, 5944–5951. doi:10.1200/jco.2008.19.6147

Pan, L., Fang, J., Chen, M.-Y., Zhai, S.-T., Zhang, B., Jiang, Z.-Y., et al. (2020). Promising Key Genes Associated with Tumor Microenvironments and Prognosis of Hepatocellular Carcinoma. *Wjg* 26, 789–803. doi:10.3748/wjg.v26.i8.789

Pan, X.-B., Lu, Y., Huang, J.-L., Long, Y., and Yao, D.-S. (2019). Prognostic Genes in the Tumor Microenvironment in Cervical Squamous Cell Carcinoma. *Aging* 11, 10154–10166. doi:10.18632/aging.102429

Piawah, S., and Venook, A. P. (2019). Targeted Therapy for Colorectal Cancer Metastases: A Review of Current Methods of Molecularly Targeted Therapy and the Use of Tumor Biomarkers in the Treatment of Metastatic Colorectal Cancer. *Cancer* 125, 4139–4147. doi:10.1002/cncr.32163

Qi, D., and Wu, E. (2019). Cancer Prognosis: Considering Tumor and its Microenvironment as a Whole. *EBioMedicine* 43, 28–29. doi:10.1016/j.ebiom.2019.04.031

Satelli, A., and Li, S. (2011). Vimentin in Cancer and its Potential as a Molecular Target for Cancer Therapy. *Cell. Mol. Life Sci.* 68, 3033–3046. doi:10.1007/s00018-011-0735-1

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303

Siegel, R. L., Miller, K. D., Fedewa, S. A., Ahnen, D. J., Meester, R. G. S., Barzi, A., et al. (2017). Colorectal Cancer Statistics, 2017. *CA: a Cancer J. clinicians* 67, 177–193. doi:10.3322/caac.21395

Siegel, R. L., Miller, K. D., Goding Sauer, A., Fedewa, S. A., Butterly, L. F., Anderson, J. C., et al. (2020). *Colorectal Cancer Statistics*. Hoboken: CA: a cancer journal for clinicians.

Sun, W., Shi, H., Yuan, Z., Xia, L., Xiang, X., Quan, X., et al. (2020). Prognostic Value of Genes and Immune Infiltration in Prostate Tumor Microenvironment. *Front. Oncol.* 10, 584055. doi:10.3389/fonc.2020.584055

Tusher, V. G., Tibshirani, R., and Chu, G. (2001). Significance Analysis of Microarrays Applied to the Ionizing Radiation Response. *Proc. Natl. Acad. Sci.* 98, 5116–5121. doi:10.1073/pnas.091062498

von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., et al. (2005). STRING: Known and Predicted Protein-Protein Associations, Integrated and Transferred across Organisms. *Nucleic Acids Res.* 33, D433–D437. doi:10.1093/nar/gki005

Wang, H., Wu, X., and Chen, Y. (2019). Stromal-Immune Score-Based Gene Signature: A Prognosis Stratification Tool in Gastric Cancer. *Front. Oncol.* 9, 1212. doi:10.3389/fonc.2019.01212

Xiao, B., Liu, L., Li, A., Xiang, C., Wang, P., Li, H., et al. (2020). Identification and Verification of Immune-Related Gene Prognostic Signature Based on ssGSEA for Osteosarcoma. *Front. Oncol.* 10, 607622. doi:10.3389/fonc.2020.607622

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Zhang, Y., Li, J.-Q., Jiang, Z.-Z., Li, L., Wu, Y., and Zheng, L. (2016). CD169 Identifies an Anti-tumour Macrophage Subpopulation in Human Hepatocellular Carcinoma. *J. Pathol.* 239, 231–241. doi:10.1002/path.4720

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 06 January 2022
doi: 10.3389/fgene.2021.793628

Characteristics of the Immune Cell Infiltration Landscape in Gastric Cancer to Assistant Immunotherapy

Chenlu Li[1†], *Jingjing Pan*[2†], *Yinyan Jiang*[3], *Yan Yu*[4], *Zhenlin Jin*[3]* and *Xupeng Chen*[1]*

[1]Department of Gastroenterology, Affiliated Yueqing Hospital, Wenzhou Medical University, Wenzhou, China, [2]Department of Laboratory Medicine, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, [3]Department of Hematopathology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, [4]Wenzhou Medical University, Wenzhou, China

Background: Gastric cancer (GC) was usually associated with poor prognosis and invalid therapeutical response to immunotherapy due to biological heterogeneity. It is urgent to screen reliable indices especially immunotherapy-associated parameters that can predict the therapeutic responses to immunotherapy of GC patients.

Methods: Gene expression profile of 854 GC patients were collected from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) datasets (GSE84433) with their corresponding clinical and somatic mutation data. Based on immune cell infiltration (ICI) levels, molecular clustering classification was performed to identify subtypes and ICI scores in GC patients. After functional enrichment analysis of subtypes, we further explored the correlation between ICI scores and Tumor Mutation Burden (TMB) and the significance in clinical immunotherapy response.

Results: Three subtypes were identified based on ICI scores with distinct immunological and prognostic characteristics. The ICI-cluster C, associated with better outcomes, was characterized by significantly higher stromal and immune scores, T lymphocytes infiltration and up-regulation of PD-L1. ICI scores were identified through using principal component analysis (PCA) and the low ICI scores were consistent with the increased TMB and the immune-activating signaling pathways. Contrarily, the high-ICI score cluster was involved in the immunosuppressive pathways, such as TGF-beta, MAPK and WNT signaling pathways, which might be responsible for poor prognosis of GC. External immunotherapy and chemotherapy cohorts validated the patients with lower ICI scores exhibited significant therapeutic responses and clinical benefits.

Conclusion: This study elucidated that ICI score could sever as an effective prognostic and predictive indicator for immunotherapy in GC. These findings indicated that the systematic assessment of tumor ICI landscapes and identification of ICI scores have crucial clinical implications and facilitate tailoring optimal immunotherapeutic strategies.

Keywords: gastric cancer, immune cell infiltration landscape, tumor microenvironment, immunotherapy, immune response

Edited by:
Eszter Lakatos,
Queen Mary University of London,
United Kingdom

Reviewed by:
Abhinita S. Mohanty,
Memorial Sloan Kettering Cancer
Center, United States
Yeqian Zhang,
Shanghai JiaoTong University, China

Correspondence:
Zhenlin Jin
jinzhenlin163@163.com
Xupeng Chen
13506551205@qq.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *12 October 2021*
Accepted: *08 December 2021*
Published: *06 January 2022*

Citation:
Li C, Pan J, Jiang Y, Yu Y, Jin Z and Chen X (2022) Characteristics of the Immune Cell Infiltration Landscape in Gastric Cancer to Assistant Immunotherapy. Front. Genet. 12:793628. doi: 10.3389/fgene.2021.793628

INTRODUCTION

As one of the most common tumors with a high morbidity and mortality, gastric cancer (GC) leads to a poor prognosis and increases critical social burden with 5.7% incidence and 8.2% mortality rates (Bray et al., 2018). More than 50% of diagnosed GC patients were at advanced stages and the prognosis of GC was relatively poor with only less than 30% overall 5-year survival rate (Yang et al., 2020a; Wang et al., 2021). Despite remarkable progress have been made for the treatment of GC, including radiotherapy, chemotherapy and surgery according to different locations and clinical stages, there is still lack of effective strategies for the advanced GC treatment (Ai and Wang, 2020). Recently, the rapid rise of immunotherapy has brought a new therapeutic landscape for the patients who didn't benefit from conventional chemotherapy, radiation or surgery (Chivu-Economescu et al., 2018). However, in clinical practice, the majority of GC patients were usually still lack of effective therapeutical response to immunotherapy (Li et al., 2020). Therefore, it is crucial to screen reliable index especially immunotherapy-related biological parameters that can predict the therapeutic responses to immunotherapy of GC patients.

Tumor microenvironment (TME) is the inner environment of malignant tumor progression and reveals the biological process of host anti-tumor immune response and destruction of normal tissue. Therefore, the TME was considered as an essential element for exploring the relationship between immune response and tumors with immune cell infiltration (ICI) (Anderson et al., 2006). The TME of tumor tissue was usually complex and associated with tumor initiation, development and prognosis, of which massive immune cells were infiltrated and played great significance to the prognosis of patients (Chen et al., 2020). For instance, tumor-infiltrating lymphocytes (TLS) such as CD4$^+$ T cell and CD8$^+$ T cell could remarkably improve the curative effects and survival rates (Vassilakopoulou et al., 2016). In addition, tumor-associated macrophages (TAMs), accounting for the majority of leukocytes, had been reported to be involved in the progress of lung and kidney tumors through secreting immunosuppressive cytokines (Vilaseca et al., 2017; Tie et al., 2020). Besides various immune cells, the hypernomic infiltration of stromal components in tumor tissues could decrease the TLS trafficking into tumors (Senbabaoglu et al., 2016). All these researches indicated that intercellular relationships were more significant than the single cell population in TME and the comprehensive landscape of immune cells infiltrating of TME in GC patients still remained unclear.

The identification of potential subtypes of GC by high-throughput technologies may contribute to elucidating the molecular mechanism, improving therapeutic response, and providing insight into any possible evaluating indicators for immunotherapy. In this study, based on the gene expression profile of GC. we applied two major computational algorithms, CIBERSORT and ESTIMATE, to acquire immune clustering subtypes, establish the immune cells infiltration (ICI) scores and further assess the comprehensive landscape about the infiltration of immune cells in GC. Besides, the biological characteristics of ICI subgroups was elucidated and the significance of ICI scores in the prediction of immunotherapy and common chemotherapeutics response was further estimated to validate the ICI landscape for GC.

Conclusively, we are convinced that this study would help in the identification of potential subtypes of GC for interpreting the discriminatory curative responses to immunotherapy and facilitating understanding of the underlying mechanisms of the disease.

MATERIALS AND METHODS

Data Preparation and Preprocessing

The flow chart of our study was showed in **Supplementary Figure S1**. Transcriptome profiling data of 854 GC samples with their corresponding clinical and mutation data were downloaded from two publicly available datasets, of which the RNA-seq transcriptome data of 407 GC patients with fragments per-kilobase million (FPKM) value were derived from The Cancer Genome Atlas (TCGA) datasets (https://portal.gdc.cancer.gov/) and other microarray data of 447 GC patients (GSE84433) were derived from the Gene Expression Omnibus (GEO) datasets (https://www.ncbi.nlm.nih.gov/geo/). To standardize the expression levels between different sequencing technologies, the FPKM value of TCGA-STAD datasets was transformed into the transcripts per-kilobase million (TPMs) form, which was consistent with the microarray datasets (Wagner et al., 2012). In addition, the "ComBat" algorithm of "sva" package was further applied to remove the non-biological technical biases due to batch effects between different datasets (Leek et al., 2012).

Consensus Cluster Analysis for Immune Cells Infiltration in Gastric Cancer

To evaluate the immune cell infiltration (ICI) characteristics of GC tissues, we used the "CIBERSORT" package (Chen et al., 2018) to quantitatively analyze the infiltration levels of different immune cells with the LM22 signatures by 1,000 random permutations. Immune cell infiltration levels and stromal contents in different samples were evaluated by using the "ESTIMATE" algorithm (Yoshihara et al., 2013). Then, we performed hierarchical clustering analysis according to the immune infiltration of each sample and the "ConsensuClusterPlus" R package (Wilkerson and Hayes, 2010) was applied to conducted unsupervised clustering based on Euclidean distance and Ward's linkage methods with 1,000 repeated times to ensure the stability of classification. We performed multiple comparisons among different immune-subtypes including tumor microenvironment (TME) and ICI features to explore the immune characteristics. Moreover, R packages "survival" (Therneau and Lumley, 2015) and "survminer" (Kassambara et al., 2017) were used to perform Kaplan-Meier survival analysis and create survival curves between immune subtypes.

Identification of ICI Gene-Types and Functional Enrichment Analysis

ICI-associated genes were identified among ICI subtypes using the "limma" package (Smyth, 2005) through setting significance cutoff as adjusted $p < 0.05$ and absolute fold-change >1 and those genes were also divided into different clusters using hierarchical clustering. In order to clarify the biological function and characteristics of gene-clusters, Gene Ontology (GO) enrichment analysis was performed by using "ClusterProfiler" package (Yu et al., 2012) and similar comparisons between gene-types were conducted including TME, ICI and survival analysis. In addition, we also compared the difference of TNM stages between ICI clusters through the chi-square test using the "ggstatsplot" R package.

Definition and Immune Characteristics of ICI Scores

Based on the unsupervised clustering of expression value of ICI-associated genes, those GC samples were redistributed into different gene-clusters and the expression values correlated with clusters were identified as the ICI gene signatures A and B respectively. Moreover, we applied the Boruta algorithm (Kursa and Rudnicki, 2010) to reduce the dimension of above ICI gene signatures and adopted principal component 1 as the signature score by performing the principal component analysis (PCA) (Zhang et al., 2020). Finally, the method similar to Gene expression grade index was applied to define the ICI score as the following formula: $\text{ICI score} = \sum \text{PC1A} - \sum \text{PC1B}$. Subsequently, the threshold of ICI scores was identified through the "surv_cutpoint" function of "survival" package and patients were separated into High- and Low-ICI groups. The software of GESA v4.0 was used for gene set enrichment analysis (GSEA) of ICI scores in gastric cancer. To estimate the discriminative capacity of ICI scores in predicting the prognosis of GCs, the "timeROC" package was applied to draw time-dependent receiver operating characteristic (ROC) curves (Blanche and Blanche, 2019).

Calculation and Analysis of Tumor Mutation Burden

To explore the relationship between TMB and ICI score, we also downloaded the mutation data of GC patients from TCGA datasets and calculated TMB scores by using the "maftool" R package (Mayakonda et al., 2018). In addition, the correlation analysis between TMB and ICI scores was performed based on Spearman correlation coefficients and combined survival analysis was further employed to clarify the prognostic value. Furthermore, comprehensive mutation analysis was conducted by "maftools" package and mutational signatures of the top 20 genes were further chosen subsequent comparison between ICI-score subgroups using chi-square test.

Exploration of the Significance of ICI Scores in Clinical Immunotherapy Response

Another independent dataset, IMvigor210, included 298 urothelial cancer patients receiving anti-PD-L1 immunotherapy with complete clinical information and was downloaded from the freely available "IMvigor210CoreBiologies" package (http://research-pub.gene.com/IMvigor210CoreBiologies). Moreover, to evaluate the therapeutic value of ICI scores in the clinic for GC treatment, we calculated the half maximal inhibitory concentration (IC50) of common chemotherapeutic drugs based on Genomics of Drug Sensitivity in Cancer (GDSC) databases (Yang et al., 2013). From the GDSC database, Antitumor drugs such as 5-Fluorouracil, Bleomycin, Cisplatin, Docetaxel and Mitomycin-C have been recommended for the GC treatment by current clinical guidelines. Difference of IC50 of these chemotherapeutic drugs between ICI-score subgroups was compared using Wilcoxon test and the results were exhibited in box diagrams using the "ggpubr" package (Whitehead et al., 2019).

RESULTS

The Landscape of Immuno-Cell Infiltration in the TME of Gastric Cancer

First, the "CIBERSORT" and "ESTIMATE" algorithm were used to calculate the activity or enrichment levels of immune cells in GC tumor tissues (**Supplementary Table S1,2**). Unsupervised clustering was applied to classify the GC patients into distinct subtypes by the "ConsesusClusterPlus" package based on 854 tumor samples with matched immune cell infiltration (ICI) profiles from the databases (GSE84433 and TCGA-GC). Three independent ICI subtypes had been identified with significant survival differences (log rank test, $p = 0.012$) and ICI analysis revealed complicated relation among immune cells in the TME of GC tissues (**Figure 1A,D**). In order to further examine the relationship of intrinsic biological differences and distinct clinical phenotypes, we compared the composition of immune cells in TME according to the three ICI subtypes. Among the three subtypes, the ICI cluster C exhibited the longer median survival time (**Figure 1E**) and higher infiltration of T lymphocytes including $CD8^+$ T cells, activated memory $CD4^+$ T cells, follicular helper T (Tfh) cells, M1 macrophages and resting dendritic cells (DCs) (**Figure 1B,C**). With a median survival of 4 years, the ICI cluster A had the worst prognosis and it was characterized by high infiltration of naive B cells, resting memory $CD4^+$ T cells, activated DCs and resting Mast cells. The ICI cluster B was marked by high infiltration of M0 and M1 macrophages with about 4.8 years' median survival time. Moreover, we also analyzed the expression of significant immune checkpoint, PD-L1, in each ICI cluster to estimate the response to immunotherapy. Consistent with the results of survival analysis, the expression of PD-L1 was higher in ICI cluster C than that in ICI cluster A and B (**Figure 1F**). In addition, the comparison of TNM stages showed that Cluster A displayed more proportion of severe stages than that of Cluster B and C (**Figure 1G**).

Identification of ICI Gene-Types and its Functional Enrichment

To further elucidate the underlying biological characteristics of different immunophenotypes, the differential transcriptome variations analysis was performed among subtypes using the

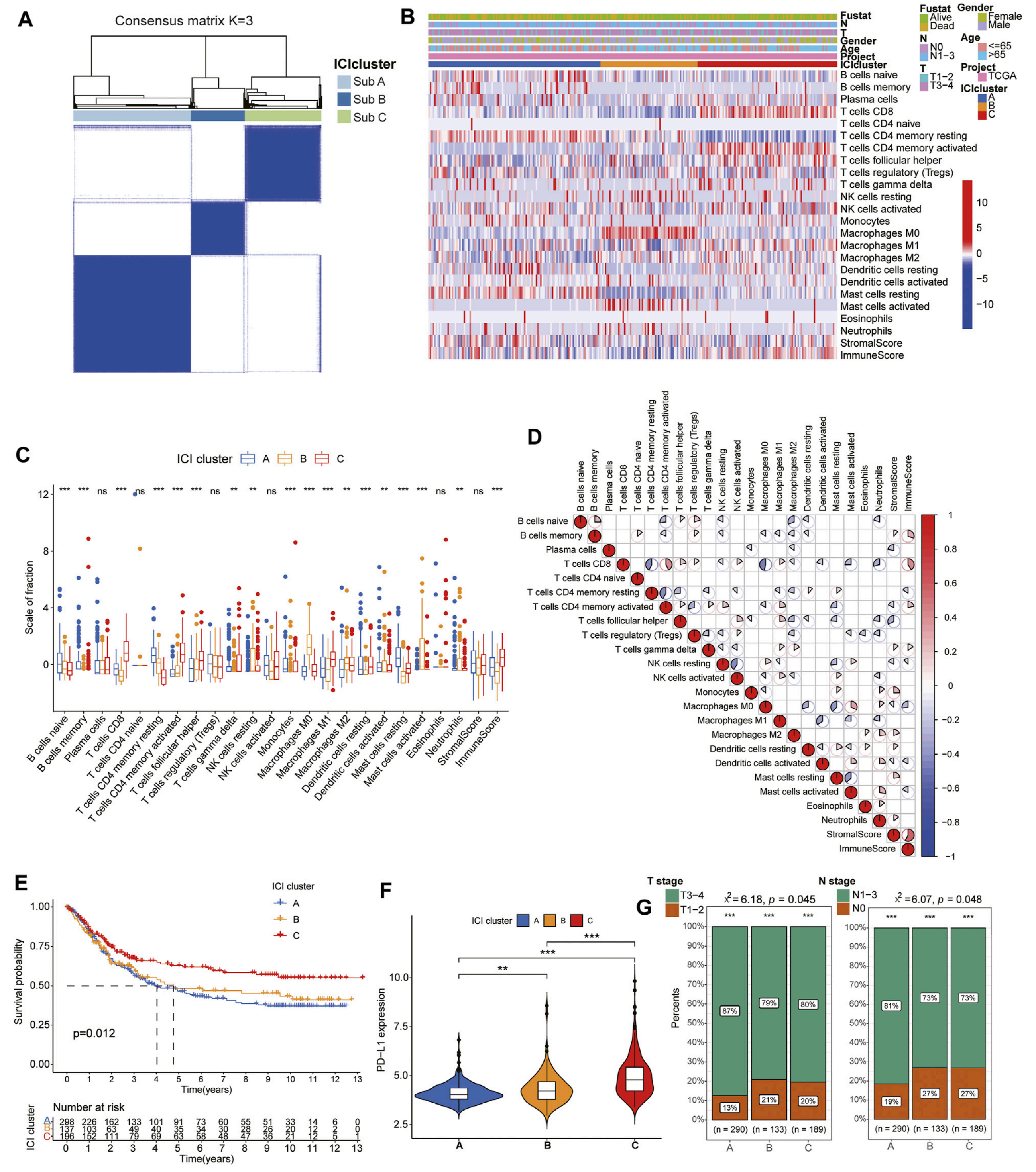


FIGURE 1 | Identification of immune molecular subtypes and characteristics of immuno-cell infiltration landscape in the gastric cancer. **(A)** Consensus clustering matrix for k = 3 in GC patients. **(B)** Heatmap of immune cells infiltration and clinicopathologic features of the three subtypes. **(C)** The box plots showing the difference of immune cells infiltration among three ICI clusters. **(D)** The correlation among the immune cell infiltration in GC patients. **(E)**. Kaplan-Meier curves of overall survival (OS) for the GC patients in three subtypes. **(F)** The expression of PD-L1 between different ICI cluster groups. **(G)** Difference of TMN stages among different ICI cluster groups.

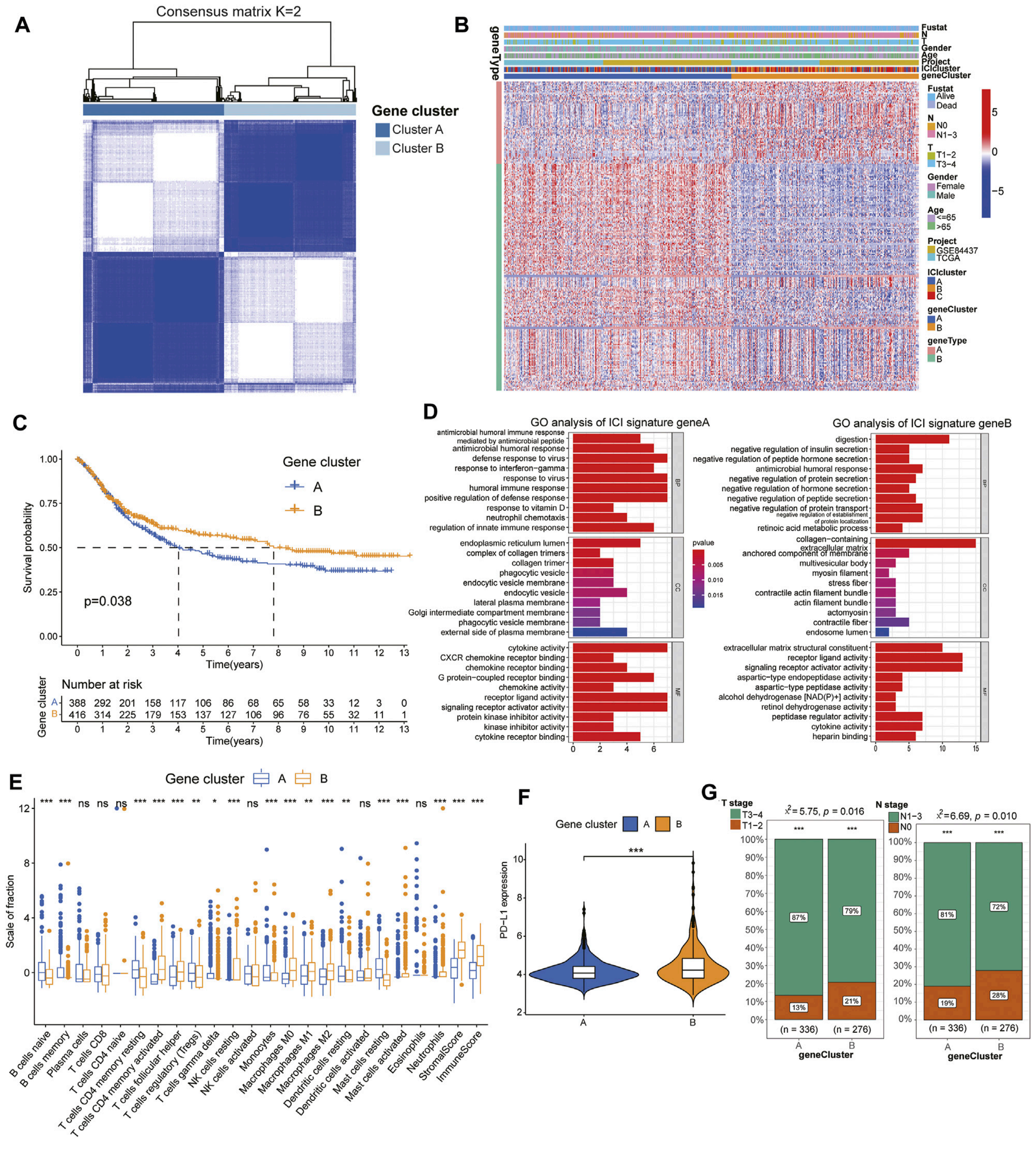


FIGURE 2 | Identification of ICI gene-types and its functional enrichment. **(A)** Consensus clustering matrix for k = 2 in GC patients based on the expression of DEGs. **(B)** Unsupervised clustering of DEGs to classify GC patients into novel two gene clusters **(A,B)**. **(C)** Kaplan-Meier curves for the two gene clusters of patients. **(D)** Gene Ontology enrichment analysis of the two ICI-related signature genes. **(E)** The difference of immune cells infiltrating in TME between two gene clusters. **(F)** The expression of PD-L1 between different gene cluster groups. **(G)** Difference of TMN stages between different gene cluster groups.

"limma" package. Subsequently, we reperformed the unsupervised hierarchical clustering based on the expression of 251 differentially expressed genes (DEGs) and classified the GC cohorts into two genomic clusters named gene clusters A and B (**Figure 2A**, **Supplementary Table S4**). Moreover, those DEGs were positively/negatively associated with ICI signatures and also classified into two clusters: ICI signature gene A and B (**Figure 2B**) and survival analysis exhibited gene clusters A

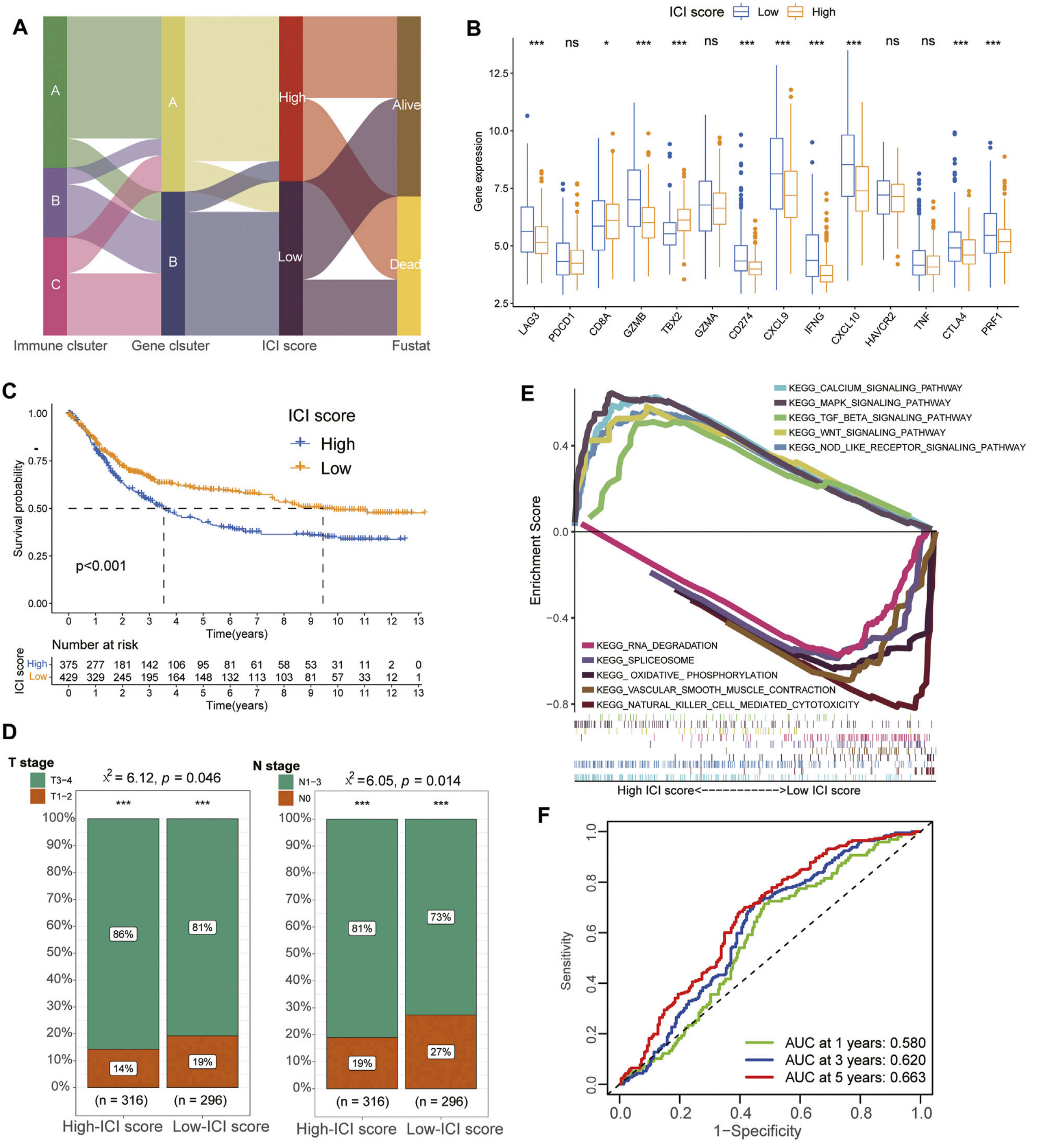


FIGURE 3 | Construction and identification of characteristics for ICI Score. **(A)** Alluvial diagram showing the ICI gene cluster distribution from different ICI gene clusters, ICI score groups and final survival outcomes. **(B)** The expression of immune-checkpoint-associated signatures (CD274/PD-L1, PDCD1, LAG3, CTLA4 and HAVCR2) and immune-activity-related genes (CD8A, CXCL9, CXCL10, GZMA, GZMB, PRF1, IFNG, TNF and TBX2) in different ICI score groups. **(C)** Kaplan-Meier curves of overall survival (OS) for the GC patients in high and low ICI score groups. **(D)** Difference of TMN stages between different ICI score groups. **(E)** The results of GSEA showing that Calcium signaling pathway, MAPK signaling pathway, TGF beta signaling pathway, WNT signaling pathway and NOD like receptor signaling pathway were significantly enriched in high-ICI score group while RNA degradation, Spliceosome, Oxidative phosphorylation, Vascular smooth muscle contraction and Natural killer cell mediated cytotoxicity were enriched in the low-ICI score group. **(F)** ROC analysis showed the 1-year, 3-year, and 5-year AUC values of the ICI scores in predicting the prognosis of GCs were 0.580, 0.620, and 0.663, respectively.

had a longer median survival time than cluster B (log rank test, $p = 0.038$, **Figure 2C**). Functional enrichment analysis revealed the ICI signature gene A was significantly enriched in the process of humoral immune response, such as antimicrobial humoral response, defense response to virus and response to interferon-gamma, while the ICI signature gene B was associated with the regulation of digestion, including negative regulation of insulin secretion, peptide hormone secretion and protein secretion (**Figure 2D**, **Supplementary Table S5**). In addition, in order to explore the prognostic implications of the ICI gene clusters, we also performed the Kaplan-Meier survival analysis and the gene cluster B had a better prognosis than cluster A (**Figure 2C**). Interestingly, TME analysis indicated gene cluster B had higher infiltration of immune cells and were associated with significantly high immune scores, suggesting its pro-tumor or anti-tumor activity (**Figure 2E**). Additionally, the two genomic clusters also showed significant differences in the expression of PD-L1 and the gene cluster B exhibited higher PD-L1 levels (**Figure 2F**). Consistent with the results of survival analysis, cluster A exhibited more severe TNM features than that of cluster B, suggesting the latter cluster might have a better prognosis and efficacious curative responses (**Figure 2G**). All these results indicated the consistency between the ICI analysis and prognostic profile in different gene clusters suggesting the scientificity and rationality of our classification method.

Construction and Identification of Characteristics for ICI Score

To acquire quantitative index of ICI landscape in GC, we defined ICI scores using principal component analysis and successfully divided the patients into High- and Low-ICI score subgroups (**Supplementary Table S7**). **Figure 3A** showed the distribution procedure of different subgroups and the gene cluster A was almost divided into High-ICI score cohorts while massive cluster B was contributed into Low-ICI score subgroups. Furthermore, we also evaluated the immune activity and immune tolerance condition of each cohort before determining the prognostic value of ICI scores. To accomplish the evaluation, immune-checkpoint-associated signatures were chosen to assess the response of immunotherapy including CD274/PD-L1, PDCD1, LAG3, CTLA4 and HAVCR2 while immune-activity-related genes were selected to estimate the condition of immune activation including CD8A, CXCL9, CXCL10, GZMA, GZMB, PRF1, IFNG, TNF and TBX2. We observed that most signatures of immune-checkpoint and immune-activity-relevant genes were significantly upregulated in the Low-ICI score groups except PDCD1, CD8A, HAVCR2, TBX2 and TNF (**Figure 3B**) and Low-ICI group also had a better prognosis than High-ICI score cohorts (**Figure 3C**). Clinical analysis of TNM stages also demonstrated that Low-ICI scores subgroups exhibited more frequent phenotypes with high-levels of clinical stages (**Figure 3D**). Moreover, GSEA analysis also revealed that Calcium signaling pathway, MAPK signaling pathway, TGF beta signaling pathway, WNT signaling pathway and NOD like receptor signaling pathway were significantly enriched in high-ICI score group while RNA degradation, Spliceosome, Oxidative phosphorylation, Vascular smooth muscle contraction and Natural killer cell mediated cytotoxicity were enriched in the low-ICI score group (**Figure 3E**, **Supplementary Table S6**). Moreover, time-dependent ROC analysis showed the 1-year, 3-year, and 5-year AUC values of the ICI scores in predicting the prognosis of GCs were 0.580, 0.620 and 0.663, respectively (**Figure 3F**).

The Relationship Between ICI Scores and Tumor Burden Mutation

Increasing evidence indicated that tumor burden mutation (TMB) could affect the infiltration of $CD8^+$ T cells, which was considered as significant elements in alleviating the prognosis of tumors (Rizvi et al., 2015; Cristescu et al., 2018). These studies implied that TMB might determine the individual's response to target immunotherapy. To investigate the intrinsic relationship between TMB and ICI scores, we compared the levels of TMB scores between ICI score subgroups and performed Spearman correlation analysis. The results revealed high-ICI score group had a lower TMB scores than that of low-ICI score cohorts (Wilcox test, $p < 0.05$) and the ICI scores were negatively correlated with TMB scores (Spearman coefficient: $R = -0.52$, $p < 2.2e\text{-}16$) (**Figure 4A,B**). Longer median survival time was also identified in high-TMB groups than low-TMB subgroups by survival analysis, consistent with the prognosis of low-ICI score groups (**Figure 4C**). Moreover, the stratified survival analysis further revealed patients with high TMB and low ICI scores had the best prognosis status, suggesting the synergistic effect of TMB and ICI scores in prognostic stratification of GC. Besides, low-ICI score cohorts still exhibited a better prognosis than that of high-ICI groups even in GC patients with same TMB stages and in patients with same ICI-score condition, high-TMB patients also had a longer median survival time than low-TMB cohorts (**Figure 4D**). These results implied the negative correlation between ICI scores and TMB values and their potential complementary value in the application of prognosis for GC patients.

Furthermore, we estimated the distribution of somatic variants between the low and high ICI subgroups based on the TCGA datasets. The results revealed various mutation patterns were identified in both high- and low-ICI subgroups including Missense Mutation, Nonsense Mutation, Frame Shift Del and In Frame Del, and more frequent mutations were observed in low-ICI groups (96.34%) than that of high-ICI groups (78.95%). The top20 genes with most mutation frequency were exhibited in **Figure 4E,F,** of which TTN, PIK3CA, KMT2D and OBSCN were significantly different between the low and high ICI score groups (chi-square test; $p < 0.05$) and the top20 genes with significantly difference were displayed in **Table1**. These results might propose novel ideas for exploring the potential mechanism of tumor ICI composition and gene mutation in immune checkpoint therapy.

The significance of ICI scores in the prediction of immunotherapy and common chemotherapeutics response.

To further explore the role of ICI scores in predicting the therapeutic benefit in GC, the patients who accepted anti-PD-L1 immunotherapy from the IMvigor210 cohort were calculated ICI

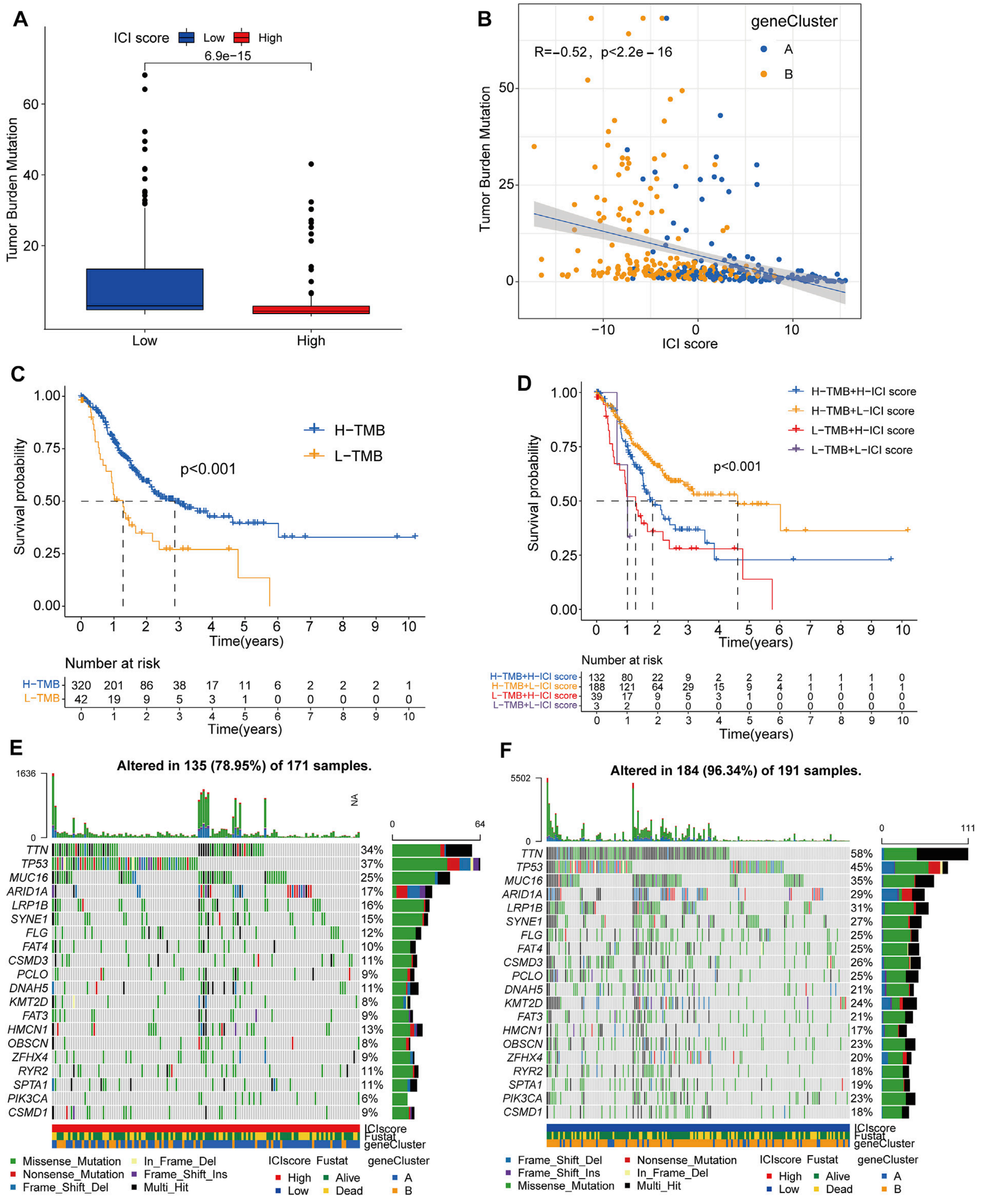


FIGURE 4 | The Relationship between ICI Scores and Tumor Burden Mutation. **(A)** The difference of TMB value between the high and low ICI score subgroups. **(B)** The scatter diagram showing the negative correlation between TMB value and ICI scores. **(C)** Kaplan-Meier curves of the high and low TMB subgroups in GC patients. **(D)** Stratified survival analysis for GC patients combining TMB groups and ICI score subtypes. **(E,F)** The oncoPrint showing the mutant situation of individual patients in high ICI scores groups (red) and low ICI scores groups (blue) respectively.

TABLE 1 | Top20 Somatic Variants between High- and Low-ICI Score group.

Gene symbol	High ICI score (%)	Low ICI score (%)	*p* Value
TTN	58 (33.92%)	111 (58.12%)	6.75E-06
PLEC	7 (4.09%)	39 (20.42%)	6.86E-06
CNTLN	3 (1.75%)	29 (15.18%)	1.65E-05
PIK3CA	11 (6.43%)	43 (22.51%)	3.48E-05
ANKRD11	4 (2.34%)	29 (15.18%)	5.00E-05
HDAC4	0 (0%)	19 (9.95%)	6.31E-05
KMT2D	13 (7.6%)	45 (23.56%)	6.64E-05
ANK3	8 (4.68%)	36 (18.85%)	7.56E-05
ASPM	4 (2.34%)	28 (14.66%)	8.25E-05
HERC2	6 (3.51%)	32 (16.75%)	8.40E-05
JARID2	1 (0.58%)	21 (10.99%)	8.91E-05
NPAP1	1 (0.58%)	21 (10.99%)	8.91E-05
SIPA1L1	2 (1.17%)	23 (12.04%)	1.11E-04
SLITRK5	4 (2.34%)	27 (14.14%)	1.35E-04
FBN1	4 (2.34%)	27 (14.14%)	1.35E-04
SSPO	4 (2.34%)	27 (14.14%)	1.35E-04
HIVEP1	1 (0.58%)	20 (10.47%)	1.49E-04
OBSCN	13 (7.6%)	43 (22.51%)	1.63E-04
KMT2A	4 (2.34%)	26 (13.61%)	2.21E-04
ATP10 A	4 (2.34%)	26 (13.61%)	2.21E-04

p *value was obtained based on the chi-square test between the high and low ICI, score subgroups.*

scores and assigned into high- or low-ICI scores groups (**Supplementary Table S8**). Notably, the effective response rate of anti-PD-L1 therapy was significantly higher in the low-ICI score group than in high-ICI subgroups and the low-ICI patients outlived the high-ICI score patients (**Figure 5A,B**). Moreover, the rate of complete remission (CR) after immunotherapy was also increased in low ICI scores than high ICI cohorts (**Figure 5C**).

Besides checkpoint blockers therapy, we also attempted to investigate the potential associations between ICI scores and the curative efficacy of common chemotherapeutics in treating gastric cancers. IC50 was calculated and five common anti-GC chemotherapeutic drugs were obtained from the GDSC databases including 5-Fluorouracil, Bleomycin, Cisplatin, Docetaxel and Mitomycin-C (**Supplementary Table S9**). Interestingly, except Mitomycin-C, other four drugs all exhibited lower IC50 value in low-ICI score groups indicating the low-ICI patients might obtain better curative efficacy from common chemotherapy (**Figures 5D–H**). Collectively, these outcomes indicated that ICI scores could be associated with the response to immunotherapy and common chemotherapy.

DISCUSSION

As a malignant tumor with high mortality, the prognosis of GC remains poor without effective therapeutical tools. Despite the development of combination chemotherapy, consisting of platinum and 5-fluorouracil, only a mild survival advantage was obtained in GC patients (Galdy et al., 2016). Recently, cancer target immunotherapies have acquired considerable attention as an effective and accurate therapeutic option for GC including immune checkpoint inhibitors, tumor vaccines and chimeric antigen rector (CAR)-T cells (Yang et al., 2019). However, even if the GC patients were at the same clinical stage, their prognosis and therapeutical response to the same treatment might be still different in clinical practice. Gullo's study has also reported this phenomenon and attributed to genomic and biological heterogeneity (Gullo et al., 2018). Therefore, identification of a novel subtype and reliable index to evaluate and predict the therapeutical response to immunotherapy for GS is urgently needed.

In this study, we first proposed an immune molecular subtype based on clustering immune infiltration scores with distinct clinical and immunological signatures in the meta-cohort of 854 GC patients. Interestingly, the characteristics of the three molecular subtypes manifested significant homogeneity. TME analysis revealed higher stromal and immune scores were found in ICI cluster C than other two clusters, indicating anti-tumor immune response was significantly activated in ICI cluster C of GC (Zeng et al., 2019). Moreover, higher infiltration scores of T cells, especially activated $CD4^+$ memory T cells, $CD8^+$ T cells and follicular helper T cells, which have been regarded as the major immune cells for anti-tumor efficacy (Biase et al., 20192019), were demonstrated in the ICI cluster C and this subtype also presented longer median survival time than other two clusters through Kaplan-Meier survival analysis (**Figure 1C,E**). Immune checkpoint genes, especially PD-L1, have been demonstrated playing significant role in immune suppression in multiple tumors and the target inhibitors have also been widely applied to immunotherapy for cancers (Kim et al., 2020a). It was worth mentioning that the expression levels of PD-L1 was significantly increased in ICI cluster C subgroups suggesting a higher level of immune exhaustion and potential better therapeutical response in GC patients.

To further explore the potential biological functional features of the ICI subtypes in GC, we fetched the differential expression genes among three subtypes and identified novel ICI gene types. Interestingly, the ICI gene cluster B exhibited a better prognosis for GC than gene cluster A and was positively associated with the expression of ICI signature A, which were significantly enriched in the process of humoral immune response. Conversely, the ICI gene cluster A was positively associated with the ICI signature B and major enriched in the negative regulation of digestion. Increasing evidence had indicated that the *H. pylori* bacteria could actively dampen the T-helper 1 (Th1) response and inhibit CD4/CD8 positive T cell activation and IFN-γ production, leading to considerable tissue damage during the progression of GC (Wen et al., 2004; Ito et al., 2008). Therefore, the process of humoral immune response would ameliorate the disease condition and improve the survival for GS patients, interpreting the better prognosis of ICI gene cluster B in our study (Kurtenkov et al., 2007). In addition, we also observed ICI gene cluster B had the higher stromal scores, immune scores, expression levels of PD-L1, milder TNM stages and other immune-response-related cells infiltration, consistent with the results of survival analysis and GO functional enrichment analysis. These outcomes suggested the ICI gene cluster B was associated with the immune-activation condition with better therapeutic reaction and prognosis for GC (Panda et al., 2018).

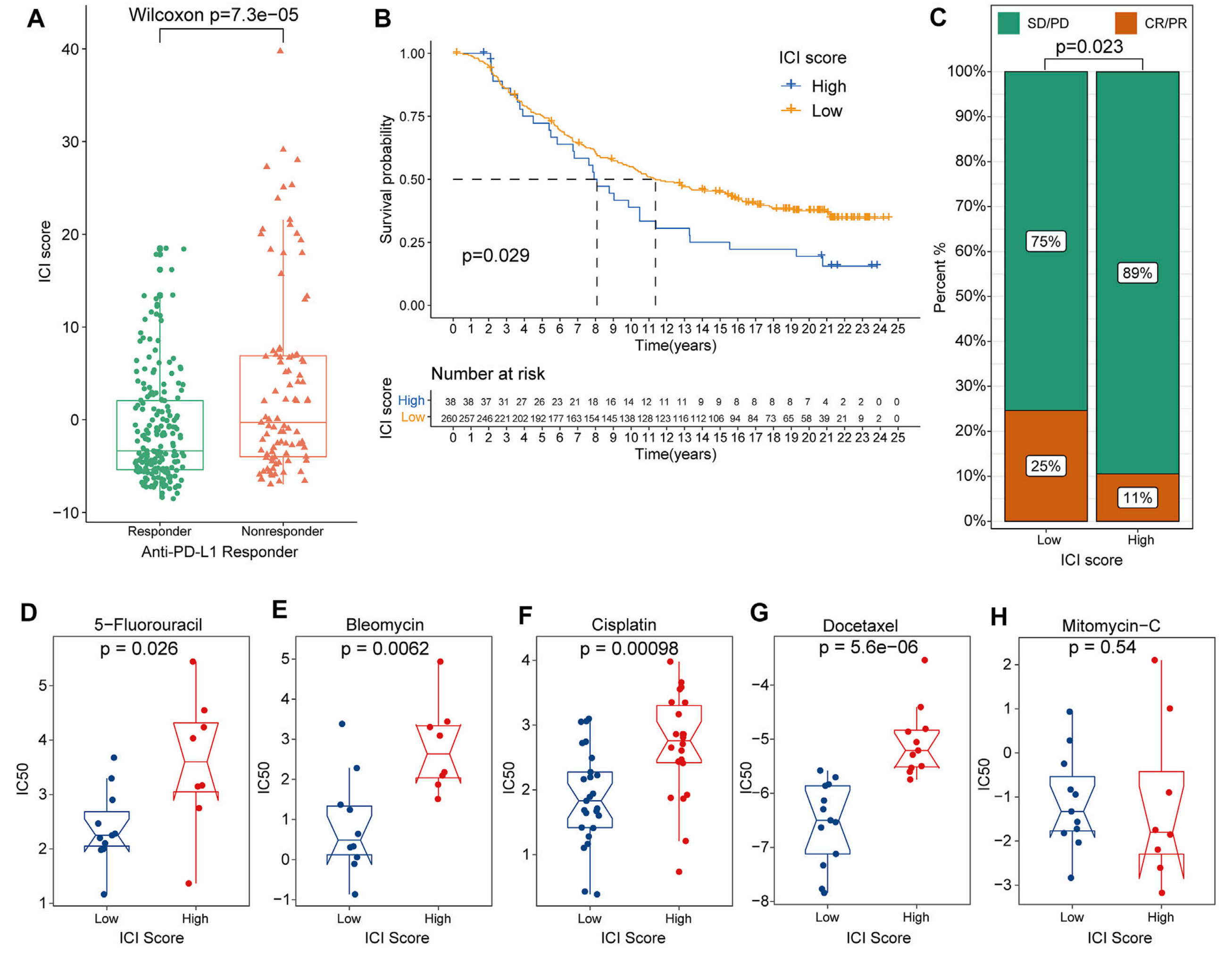


FIGURE 5 | The role of ICI scores in the prediction of immunotherapy and common chemotherapeutics response. **(A)** ICI scores between groups with different clinical immunotherapy response status. **(B)** Survival analysis for patients in high and low ICI score groups from the IMvigor210 cohort. **(C)** The distribution of the complete remission (CR)/partial response (PR) rate and stable disease (SD)/progressive disease (PD) to anti-PD-L1 immunotherapy between high and low ICI score groups based on the IMvigor210 cohort. **(D–H)** The difference of IC50 value from five common chemotherapy drugs between high and low ICI score groups, including 5-Fluorouracil, Bleomycin, Cisplatin, Docetaxel and Mitomycin-C.

Considering the individual biological heterogeneity to immune checkpoint inhibitors, it was urgently required to understand the ICI landscape of individual tumors.

In previous studies, tumor subtype-specific biomarkers had been successfully established to improve individual outcome prediction in breast and colorectal cancers, respectively (Callari et al., 2016; Bramsen et al., 2017). In our study, through the Boruta algorithm, we successfully established the ICI score to quantify the ICI pattern and found most low-ICI score groups were corresponding to the former ICI gene cluster B with a longer lifetime. Moreover, the expression levels of most immune-checkpoint-associated and immune-activity-related genes were both significantly increased in the low-ICI score groups, implying the activation of anti-tumor process in gastric cancers. ROC analysis further demonstrated that ICI scores had a good prediction capacity in all 1-year, 3-year and 5-year overall survival for GC patients, indicating the potential predicted value of ICI scores. In addition, GSEA revealed that the genes of high-ICI score cluster were involved in the immunosuppressive pathways, such as TGF-beta, MAPK and WNT signaling pathways, which had been reported associated with the progression of GC (Chen et al., 2014; Jia et al., 2017; Yang et al., 2020b). Contrarily, several immune-activated and metabolic-related pathways were found enriched in low-ICI score cohorts including Natural killer cell mediated cytotoxicity and Oxidative phosphorylation. Su et al. (2020) also identified three oxidative phosphorylation genes associated with the clinical prognostic significance in GC and multiple therapeutic technologies had been found to activate NK cells directly or indirectly to improve their killing activity for GC including cytokines, antibodies, immunomodulatory drugs, immune checkpoint blockades and gene therapy (Mimura et al., 2014).

TMB has been recognized as a new biomarker for immune checkpoint treatment in various tumor types and reported to applied in predicting the survival status after immunotherapy in advanced gastric cancer patients (Kim et al., 2020b). Therefore, TMB value was considered as a sensitive index to immunotherapy. In the current study, we also detected that the TMB was significantly increased in patients with low ICI scores. The significantly negative correlation between the TMB value and ICI scores was identified with the correlation coefficient of −0.52. The stratified analysis revealed that the prognosis value of ICI scores was consistent with TMB values and the patients with high-TMB and low-ICI scores exhibited optimal survival condition. Notably, ICI scores could still exhibit significant discriminating capacity in estimating the survival period of GC patients in same TMB conditions using stratified analysis, indicating that ICI scores might provide insights not available from TMB. Recently, the correlation between gene mutations and response or tolerance to immunotherapy had been identified in published reports (George et al., 2017). In our study, more frequent mutations were observed in low-ICI groups and massive mutable genes with significant variant frequency differences were identified, suggesting somatic mutation might participate in the process of immune-subtypes in GC.

Furthermore, to validate the significance of ICI scores in the prediction of immunotherapy, the patients receiving immunotherapy were evaluated based on the IMvigor210 datasets and we found the ICI score was significantly decreased in patients responded to corresponding immunotherapy, suggesting target immunotherapy might be beneficial tool for the patients with low ICI scores. In addition, the low-ICI score groups also demonstrated longer median survival time and higher rate of complete remission (CR) after immunotherapy in clinical trials. Besides immunotherapy, common chemotherapeutic drugs also be demonstrated lower IC50 value in low-ICI score cohorts including 5-Fluorouracil, Bleomycin, Cisplatin and Docetaxel from GDSC database, implying the low-ICI score patients might be more efficacious against these chemotherapeutic drugs. Overall, these findings from external datasets validated the potential benefits in low-ICI scores and indicated ICI scores might play a vital role in predicting the curative responses to common chemotherapy and immune checkpoint therapy.

However, there are still several limitations in our study. For one thing, the high-throughput sequencing datasets for initial analysis were relatively insufficient because it was simply obtained from the public databases. The corresponding results and conclusion remain to be investigated through more external congeneric researches. For another, there are still several concerns need other researches, even clinical practices, to repeatedly confirm and improve, such as the concrete role of ICI scores in predicting the response to immunotherapy, the optimal threshold for the classification ICI scores and so on.

CONCLUSION

In conclusion, we comprehensively explored the ICI landscape of GC, providing a clear visual angle of the characteristics in immune molecular subtypes based on clustering immune infiltration scores with distinct clinical and immunological signatures. The distinction in ICI landscapes was found to be associated with the complexity and heterogeneity of tumor treatment. Moreover, we successfully identified and validated the significance of ICI scores in predicting the therapeutic responses to immunotherapy based on clinical trial data from external datasets. The systematic assessment of tumor ICI landscapes and identification of ICI scores have crucial clinical implications and facilitate tailoring optimal immunotherapeutic strategies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

CL and JP contributed to data analysis and drafting of the manuscript. YJ and YY contributed to data acquisition. ZJ contributed to figures presentation and revision of the manuscript. XC contributed to the design of the study. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.793628/full#supplementary-material

REFERENCES

Ai, L., and Wang, H. (2020). Effects of Propofol and Sevoflurane on Tumor Killing Activity of Peripheral Blood Natural Killer Cells in Patients with Gastric Cancer. *J. Int. Med. Res.* 48, 300060520904861. doi:10.1177/0300060520904861

Anderson, A. R. A., Weaver, A. M., Cummings, P. T., and Quaranta, V. (2006). Tumor Morphology and Phenotypic Evolution Driven by Selective Pressure from the Microenvironment. *Cell* 127, 905–915. doi:10.1016/j.cell.2006.09.042

Biase, S. D., Ma, X., Wang, X., Yu, J., Wang, Y.-C., Smith, D. J., et al. (20192019). Creatine Uptake Regulates CD8 T Cell Antitumor Immunity. *jem* 216, 2869–2882. doi:10.1084/jem.20182044

Blanche, P., and Blanche, M. P. (2019). *Package 'timeROC'*.

Bramsen, J. B., Rasmussen, M. H., Ongen, H., Mattesen, T. B., Ørntoft, M.-B. W., Árnadóttir, S. S., et al. (2017). Molecular-Subtype-Specific Biomarkers Improve Prediction of Prognosis in Colorectal Cancer. *Cell Rep.* 19, 1268–1280. doi:10.1016/j.celrep.2017.04.045

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J. Clinicians* 68, 394–424. doi:10.3322/caac.21492

Callari, M., Cappelletti, V., D'Aiuto, F., Musella, V., Lembo, A., Petel, F., et al. (2016). Subtype-Specific Metagene-Based Prediction of Outcome after Neoadjuvant and Adjuvant Treatment in Breast Cancer. *Clin. Cancer Res.* 22, 337–345. doi:10.1158/1078-0432.ccr-15-0757

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12

Chen, F., Zhuang, M., Peng, J., Wang, X., Huang, T., Li, S., et al. (2014). Baicalein Inhibits Migration and Invasion of Gastric Cancer Cells through Suppression of the TGF-β Signaling Pathway. *Mol. Med. Rep.* 10, 1999–2003. doi:10.3892/mmr.2014.2452

Chen, W., Dai, X., Chen, Y., Tian, F., Zhang, Y., Zhang, Q., et al. (2020). Significance of STAT3 in Immune Infiltration and Drug Response in Cancer. *Biomolecules* 10, 834. doi:10.3390/biom10060834

Chivu-Economescu, M., Matei, L., Necula, L. G., Dragu, D. L., Bleotu, C., and Diaconu, C. C. (2018). New Therapeutic Options Opened by the Molecular Classification of Gastric Cancer. *Wjg* 24, 1942–1961. doi:10.3748/wjg.v24.i18.1942

Cristescu, R., Mogg, R., Ayers, M., Albright, A., Murphy, E., Yearley, J., et al. (2018). Pan-tumor Genomic Biomarkers for PD-1 Checkpoint Blockade-Based Immunotherapy. *Science* 362, eaar3593. doi:10.1126/science.aar3593

Galdy, S., Cella, C. A., Spada, F., Murgioni, S., Frezza, A. M., Ravenda, S. P., et al. (2016). Systemic Therapy beyond First-Line in Advanced Gastric Cancer: An Overview of the Main Randomized Clinical Trials. *Crit. Rev. Oncology/Hematology* 99, 1–12. doi:10.1016/j.critrevonc.2015.09.004

George, S., Miao, D., Demetri, G. D., Adeegbe, D., Rodig, S. J., Shukla, S., et al. (2017). Loss of PTEN Is Associated with Resistance to Anti-PD-1 Checkpoint Blockade Therapy in Metastatic Uterine Leiomyosarcoma. *Immunity* 46, 197–204. doi:10.1016/j.immuni.2017.02.001

Gullo, I., Carneiro, F., Oliveira, C., and Almeida, G. M. (2018). Heterogeneity in Gastric Cancer: From Pure Morphology to Molecular Classifications. *Pathobiology* 85, 50–63. doi:10.1159/000473881

Ito, T., Kobayashi, D., Uchida, K., Takemura, T., Nagaoka, S., Kobayashi, I., et al. (2008). *Helicobacter pylori* Invades the Gastric Mucosa and Translocates to the Gastric Lymph Nodes. *Lab. Invest.* 88, 664–681. doi:10.1038/labinvest.2008.33

Jia, S., Lu, J., Lu, J., Qu, T., Feng, Y., Wang, X., et al. (2017). MAGI1 Inhibits Migration and Invasion via Blocking MAPK/ERK Signaling Pathway in Gastric Cancer. *Chin. J. Cancer Res.* 29, 25–35. doi:10.21147/j.issn.1000-9604.2017.01.04

Kassambara, A., Kosinski, M., Biecek, P., and Fabian, S. (2017). Package 'survminer'[J]. Drawing Survival Curves using 'ggplot2' (R package version 03 1).

Kim, H. C., Choi, C. M. J. T., and Diseases, R. (2020). Current Status of Immunotherapy for Lung Cancer and Future Perspectives. *Tuberc. Respir. Dis. (Seoul)* 83, 14–19. doi:10.4046/trd.2019.0039

Kim, J., Kim, B., Kang, S. Y., Heo, Y. J., Park, S. H., Kim, S. T., et al. (2020). Tumor Mutational Burden Determined by Panel Sequencing Predicts Survival after Immunotherapy in Patients with Advanced Gastric Cancer. *Front. Oncol.* 10, 314. doi:10.3389/fonc.2020.00314

Kursa, M. B., and Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *J. Stat. Softw.* 36, 1–13. doi:10.18637/jss.v036.i11

Kurtenkov, O., Klaamas, K., Mensdorff-Pouilly, S., Miljukhina, L., Shljapnikova, L., and Chužmarov, V. (2007). Humoral Immune Response to MUC1 and to the Thomsen-Friedenreich (TF) Glycotope in Patients with Gastric Cancer: Relation to Survival. *Acta Oncologica* 46, 316–323. doi:10.1080/02841860601055441

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments. *Bioinformatics* 28, 882–883. doi:10.1093/bioinformatics/bts034

Li, Y., Liu, Y., Yao, J., Li, R., and Fan, X. (2020). Downregulation of miR-484 Is Associated with Poor Prognosis and Tumor Progression of Gastric Cancer. *Diagn. Pathol.* 15, 25. doi:10.1186/s13000-020-00946-8

Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and Comprehensive Analysis of Somatic Variants in Cancer. *Genome Res.* 28, 1747–1756. doi:10.1101/gr.239244.118

Mimura, K., Kamiya, T., Shiraishi, K., Kua, L.-F., Shabbir, A., So, J., et al. (2014). Therapeutic Potential of Highly Cytotoxic Natural Killer Cells for Gastric Cancer. *Int. J. Cancer* 135, 1390–1398. doi:10.1002/ijc.28780

Panda, A., Mehnert, J. M., Hirshfield, K. M., Riedlinger, G., Damare, S., Saunders, T., et al. (2018). Immune Activation and Benefit from Avelumab in EBV-Positive Gastric Cancer. *J. Natl. Cancer Inst.* 110, 316–320. doi:10.1093/jnci/djx213

Rizvi, N. A., Hellmann, M. D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J. J., et al. (2015). Cancer Immunology. Mutational Landscape Determines Sensitivity to PD-1 Blockade in Non-small Cell Lung Cancer. *Science* 348, 124–128. doi:10.1126/science.aaa1348

Senbabaoglu, Y., Gejman, R. S., Winer, A. G., Liu, M., Van Allen, E. M., de Velasco, G., et al. (2016). Tumor Immune Microenvironment Characterization in clear Cell Renal Cell Carcinoma Identifies Prognostic and Immunotherapeutically Relevant Messenger RNA Signatures. *Genome Biol.* 17, 231. doi:10.1186/s13059-016-1092-z

Smyth, G. K. (2005). *Limma: Linear Models for Microarray Data. Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Berlin: Springer, 397–420.

Su, F., Zhou, F. f., Zhang, T., Wang, D. w., Zhao, D., Hou, X. m., et al. (2020). Quantitative Proteomics Identified 3 Oxidative Phosphorylation Genes with Clinical Prognostic Significance in Gastric Cancer. *J. Cell Mol Med* 24, 10842–10854. doi:10.1111/jcmm.15712

Therneau, T. M., and Lumley, T. (2015). Package 'survival'[J]. *R Top Doc.* 128 (10), 28–33.

Tie, Y., Zheng, H., He, Z., Yang, J., Shao, B., Liu, L., et al. (2020). Targeting Folate Receptor β Positive Tumor-Associated Macrophages in Lung Cancer with a Folate-Modified Liposomal Complex. *Sig Transduct Target. Ther.* 5, 6. doi:10.1038/s41392-020-0115-0

Vassilakopoulou, M., Avgeris, M., Velcheti, V., Kotoula, V., Rampias, T., Chatzopoulos, K., et al. (2016). Evaluation of PD-L1 Expression and Associated Tumor-Infiltrating Lymphocytes in Laryngeal Squamous Cell Carcinoma. *Clin. Cancer Res.* 22, 704–713. doi:10.1158/1078-0432.ccr-15-1543

Vilaseca, A., Campillo, N., Torres, M., Musquera, M., Gozal, D., Montserrat, J. M., et al. (2017). Intermittent Hypoxia Increases Kidney Tumor Vascularization in a Murine Model of Sleep Apnea. *PLoS One* 12, e0179444. doi:10.1371/journal.pone.0179444

Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA Abundance Using RNA-Seq Data: RPKM Measure Is Inconsistent Among Samples. *Theor. Biosci.* 131, 281–285. doi:10.1007/s12064-012-0162-3

Wang, D. P., Zhao, R., Qi, Y. H., Shen, J., Hou, J. Y., Wang, M. Y., et al. (2021). High Expression of Interleukin-2 Receptor Subunit Gamma Reveals Poor Prognosis in Human Gastric Cancer. *J. Oncol.* 2021, 6670834. doi:10.1155/2021/6670834

Wen, S., Felley, C. P., Bouzourene, H., Reimers, M., Michetti, P., and Pan-Hammarström, Q. (2004). Inflammatory Gene Profiles in Gastric Mucosa duringHelicobacter pyloriInfection in Humans. *J. Immunol.* 172, 2595–2606. doi:10.4049/jimmunol.172.4.2595

Whitehead, M. J., McCanney, G. A., Willison, H. J., and Barnett, S. C. (2019). MyelinJ: an ImageJ Macro for High Throughput Analysis of Myelinating Cultures. *Bioinformatics* 35, 4528–4530. doi:10.1093/bioinformatics/btz403

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26, 1572–1573. doi:10.1093/bioinformatics/btq170

Yang, B., Bai, Q., Chen, H., Su, K., and Gao, C. (2020). LINC00665 Induces Gastric Cancer Progression through Activating Wnt Signaling Pathway. *J. Cell Biochem* 121, 2268–2276. doi:10.1002/jcb.29449

Yang, L., Wang, Y., and Wang, H. (2019). Use of Immunotherapy in the Treatment of Gastric Cancer. *Oncol. Lett.* 18, 5681–5690. doi:10.3892/ol.2019.10935

Yang, L., Wang, M., and He, P. (2020). LncRNA NEAT1 Promotes the Progression of Gastric Cancer through Modifying the miR-1224-5p/RSF1 Signaling Axis. *Cmar* 12, 11845–11855. doi:10.2147/cmar.s267666

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a Resource for Therapeutic Biomarker Discovery in Cancer Cells. *Nucleic Acids Res.* 41, D955–D961. doi:10.1093/nar/gks1111

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

Zeng, D., Li, M., Zhou, R., Zhang, J., Sun, H., Shi, M., et al. (2019). Tumor Microenvironment Characterization in Gastric Cancer Identifies Prognostic and Immunotherapeutically Relevant Gene Signatures. *Cancer Immunol. Res.* 7, 737–750. doi:10.1158/2326-6066.cir-18-0436

Zhang, X., Shi, M., Chen, T., and Zhang, B. (2020). Characterization of the Immune Cell Infiltration Landscape in Head and Neck Squamous Cell Carcinoma to Aid Immunotherapy. *Mol. Ther. - Nucleic Acids* 22, 298–309. doi:10.1016/j.omtn.2020.08.030

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 31 January 2022
doi: 10.3389/fgene.2021.819520

Prognostic Biomarker DDOST and Its Correlation With Immune Infiltrates in Hepatocellular Carcinoma

Changyu Zhu[1], *Hua Xiao*[2,3], *Xiaolei Jiang*[4], *Rongsheng Tong*[2,3]* and *Jianmei Guan*[5]*

[1]Department of Pharmacy, Sichuan Academy of Medical Science and Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China, [2]Personalized Drug Therapy Key Laboratory of Sichuan Province, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, Chengdu, China, [3]Department of Pharmacy, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, Chengdu, China, [4]Department of Pharmacy, Gansu Provincial Hospital of TCM, Lanzhou, China, [5]Central Sterile Supply Department, Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, Chengdu, China

Background: Dolichyl-diphosphooligosaccharide–protein glycosyltransferase non-catalytic subunit (DDOST) is an important enzyme in the process of high-mannose oligosaccharide transferring in cells. Increasing DDOST expression is associated with impairing liver function and the increase of hepatic fibrosis degrees, hence exacerbating the liver injury. However, the relation between DDOST and hepatocellular carcinoma (HCC) has not been revealed yet.

Method: In this study, we evaluated the prognostic value of DDOST in HCC based on data from The Cancer Genome Atlas (TCGA) database. The relationship between DDOST expression and clinical-pathologic features was evaluated by logistic regression, the Wilcoxon signed-rank test, and Kruskal–Wallis test. Prognosis-related factors of HCC including DDOST were evaluated by univariate and multivariate Cox regression and the Kaplan–Meier method. DDOST-related key pathways were identified by gene set enrichment analysis (GSEA). The correlations between DDOST and cancer immune infiltrates were investigated by the single-sample gene set enrichment analysis (ssGSEA) of TCGA data.

Results: High DDOST expression was associated with poorer overall survival and disease-specific survival of HCC patients. GSEA suggested that DDOST is closely correlated with cell cycle and immune response *via* the PPAR signaling pathway. ssGSEA indicated that DDOST expression was positively correlated with the infiltrating levels of Th2 cells and negatively correlated with the infiltration levels of cytotoxic cells.

Conclusion: All those findings indicated that DDOST was correlated with prognosis and immune infiltration in HCC.

Keywords: DDOST, HCC, tumor-infiltration, prognosis, T helper cells

Edited by:
Luis Zapata,
Institute of Cancer Research (ICR),
United Kingdom

Reviewed by:
Yi Shi,
Shanghai Jiao Tong University, China
Eleonora Lusito,
San Raffaele Telethon Institute for
Gene Therapy (SR-Tiget), Italy

Correspondence:
Rongsheng Tong
tongrs@126.com
Jianmei Guan
781067606@qq.com

Specialty section:
This article was submitted to
Cancer Genetics and Oncogenomics,
a section of the journal
Frontiers in Genetics

Received: *21 November 2021*
Accepted: *23 December 2021*
Published: *31 January 2022*

Citation:
Zhu C, Xiao H, Jiang X, Tong R and Guan J (2022) Prognostic Biomarker DDOST and Its Correlation With Immune Infiltrates in Hepatocellular Carcinoma. Front. Genet. 12:819520. doi: 10.3389/fgene.2021.819520

INTRODUCTION

As the principal histologic type of liver cancer, hepatocellular carcinoma (HCC) ranks third among the leading cause of cancer-related mortalities (Forner et al., 2018). It has been reported that the highest incidence rates of HCC cases occur in Asia and Africa (Petrick et al., 2020), where the exposure to chronic hepatitis B is the main risk factor (Howell et al., 2021). Therapeutic options for the treatment of HCC have substantially evolved over the past 10 years. Nowadays, patients diagnosed with HCC at any stage of the disease can benefit from effective treatment, which greatly improves their survival rate. However, there are still several areas that need urgent improvement. The molecular mechanisms underlying tumorigenesis and the progression of HCC remain poorly understood (Zhao et al., 2019). At present, serum alpha-fetoprotein (AFP), ultrasonography, and CT scanning are still important means for the early diagnosis of HCC; however, the misdiagnosis rate is high (Kelley et al., 2020; Singal et al., 2020). Despite intensive research, the 5-year survival rate of HCC is still as low as <12% due to the lack of early detection strategy and effective therapy (Hlady et al., 2019). As a result, the investigation of effective prognostic biomarkers is a pivotal area among several considerations within the research of HCC.

DDOST encodes dolichyl-diphosphooligosaccharide–protein glycosyltransferase non-catalytic subunit that forms oligosaccharyltransferase (OST) complex, which catalyzes high-mannose oligosaccharides transferring to asparagine residues on nascent polypeptides in the lumen of the rough endoplasmic reticulum (ER) (Yamagata et al., 1997). A short cytosolic tail of DDOST has a functional ER-retention di-lysine motif, serving as a mechanism for retaining OST in the ER (Fu and Kreibich, 2000). DDOST also played a role in the processing of advanced glycation end products (AGEs), which are formed from non-enzymatic reactions between lipids or protein and sugars and are associated with aging and many diseases including the congenital disorders of glycosylation (Jones et al., 2012; Zhuang et al., 2017). AGEs and their receptor had been proven to be upregulated in liver fibrosis, and the silencing receptor of AGEs reduced collagen deposition and the tumor growth of HCC (Hollenbach, 2017). A previous study also revealed that the

TABLE 1 | Demographic and clinicopathological parameters of high and low DDOST expression group patients with hepatocellular carcinoma in TCGA-LIHC.

Characteristic	Low expression of DDOST	High expression of DDOST	p
N	187	187	
T stage, n (%)			0.074
T1	104 (28%)	79 (21.3%)	
T2	40 (10.8%)	55 (14.8%)	
T3	37 (10%)	43 (11.6%)	
T4	5 (1.3%)	8 (2.2%)	
N stage, n (%)			0.622
N0	128 (49.6%)	126 (48.8%)	
N1	1 (0.4%)	3 (1.2%)	
M stage, n (%)			1.000
M0	133 (48.9%)	135 (49.6%)	
M1	2 (0.7%)	2 (0.7%)	
Pathologic stage, n (%)			0.129
Stage I	99 (28.3%)	74 (21.1%)	
Stage II	40 (11.4%)	47 (13.4%)	
Stage III	37 (10.6%)	48 (13.7%)	
Stage IV	3 (0.9%)	2 (0.6%)	
Gender, n (%)			0.269
Female	55 (14.7%)	66 (17.6%)	
Male	132 (35.3%)	121 (32.4%)	
OS event, n (%)			0.013
Alive	134 (35.8%)	110 (29.4%)	
Dead	53 (14.2%)	77 (20.6%)	
Vascular invasion, n (%)			0.072
No	118 (37.1%)	90 (28.3%)	
Yes	50 (15.7%)	60 (18.9%)	
Histologic grade, n (%)			0.077
G1	33 (8.9%)	22 (6%)	
G2	96 (26%)	82 (22.2%)	
G3	52 (14.1%)	72 (19.5%)	
G4	5 (1.4%)	7 (1.9%)	
Age, median (IQR)	62 (53, 69)	60.5 (51, 68)	0.306
AFP(ng/ml), median (IQR)	7.5 (3, 113.5)	27.5 (7, 738.75)	<0.001
BMI, median (IQR)	25.03 (22.42, 29.65)	24.16 (20.96, 28.03)	0.031

increased expression of DDOST was significantly associated with poorer clinical outcomes in cutaneous squamous cell carcinoma (Shapanis et al., 2021). Based on the those research, DDOST may play an important role in HCC. However, the prognostic potential of DDOST for HCC has not been reported.

MATERIALS AND METHODS

Data Acquisition and Preprocessing

We utilized The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/) for liver hepatocellular carcinoma (LIHC) to obtain the RNA-Seq data of 374 HCC patients accompanied with 50 normal tissues on gene expression, immune system infiltrates, and related patients' clinical information (Blum et al., 2018). Then, we transferred RNAseq data in FPKM format to TPM format, retained clinical data and RNAseq data, and further analyzed all data in accordance with the publication guidelines provided by TCGA (https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga/using-tcga).

Differentially Expressed Gene Analysis

The expression data (HTseq-Counts) were divided into high and low expression groups according to the median DDOST expression level and was then further analyzed by unpaired Student's t-test within the DESeq2 R package (3.6.3) (Love et al., 2014). Adjusted p <0.05 and |log2-fold change (FC)| >1.5 were considered as thresholds for the DEGs.

Enrichment Analysis

Gene ontology (GO) functional enrichment analysis and gene set enrichment analysis (GSEA) were all performed by ClusterProfiler package in R (3.6.3) (Yu et al., 2012). The DEGs between the high and low expression levels of DDOST were selected to be analyzed. GO analysis includes cellular component (CC), molecular function (MF), and biological process (BP). GSEA is a computational method to determine whether an *a priori* defined set of genes has statistical significance and concordant differences in two biological states. Additionally, the normalized enrichment score (NES) and adjusted p-value were utilized to sort the enriched pathways in each phenotype (Subramanian et al., 2005). C2. Cp.v7.2. symbols.gmt [Curated] was selected as the reference gene set of the KEGG pathway, C5. All.v7.2. symbols.gmt [Gene ontology] was selected as the reference gene set of GO term. Gene sets with a false discovery rate (FDR) <0.25 and adjusted p <0.05 were considered significantly enriched.

Immune Infiltration Analysis

ssGSEA was realized by the GSVA package (Hänzelmann et al., 2013) in R to investigate the correlation between DDOST and the signature genes of 24 types of immune cells and then systematically analyzed the immune infiltrates of DDOST in the published literature (Bindea et al., 2013). The infiltration of immunocytes between the DDOST high and low expression group was analyzed by Spearman correlation and the Wilcoxon rank-sum test.

Protein–Protein Interaction Network

The protein–protein interaction (PPI) network of co-regulated DEGs and the functional interaction between proteins were analyzed by the Search Tool for the Retrieval of Interacting Genes database (http://string-db.org) (Szklarczyk et al., 2019) and visualized by Cytoscape software (version 3.7.2). The combined score threshold of interaction in our study was 0.7. The database has a comprehensive score for each pair of protein relationships distributed between 0 and 1; the higher the total score, the more reliable the PPI relationship.

Validation Analysis

The different DDOST expressions between HCC and non-tumor tissue was also analyzed in three RNAseq datasets (GSE87630, GSE101685, and GSE60502), which were downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo).

The Kaplan–Meier (K-M) plotter is capable to assess the effect of 54 k genes on survival in 21 cancer types (http://kmplot.com/analysis/index.php?p=service&cancer=liver_rnaseq). The sources for the databases include GEO, EGA, and TCGA. The primary purpose of the tool is a meta-analysis-based discovery and validation of survival biomarkers (Menyhárt et al., 2018). DDOST was inputted in the K-M plotter to analyze the relationship between the expression of DDOST and the survival days of HCC patients, which were visualized in K-M survival plots. The log rank p-value <0.05 was considered statistically significant.

Statistical Analysis

The statistical data acquired from TCGA were processed by R 3.6.3. The expression levels of DDOST between HCC and the normal group were compared by Wilcoxon rank-sum test and Wilcoxon signed-rank test. The correlation between DDOST expression and the grade of clinicopathological factors was analyzed by Welch one-way ANOVA, followed by the Bonferroni correction or *t*-test. The effect of the clinicopathological factors on DDOST expression was analyzed by univariate logistic regression, the Fisher exact test, and normal and adjusted Pearson κ^2 tests. Moreover, we combined univariate Cox regression analysis and multivariate Cox regression analysis to evaluate the prognostic value of DDOST expression and other clinicopathological factors on overall survival (OS). All variables in the univariate analysis were put into the multivariate analysis. The K-M curve was drawn to evaluate the prognostic value of DDOST. The hazard risk (HR) of the individual for OS and disease-specific survival (DSS) were estimated by univariate Cox proportional hazard regressions. The HR of individual factors was estimated by measuring the HR with a 95% confidence interval (CI).

The receiver operating characteristic (ROC) analysis of DDOST was realized by the pROC package (Robin et al., 2011). The calculated area under the curve (AUC) value ranges, which were from 0.5 to 1.0, indicated the discrimination ability of 50%–100%. The time-dependent analysis of the ROC curve was constructed to evaluate DDOST for predicting the HCC outcome at 1, 3, and 5 years. All statistical tests were considered significant when two-tailed p ≤0.05.

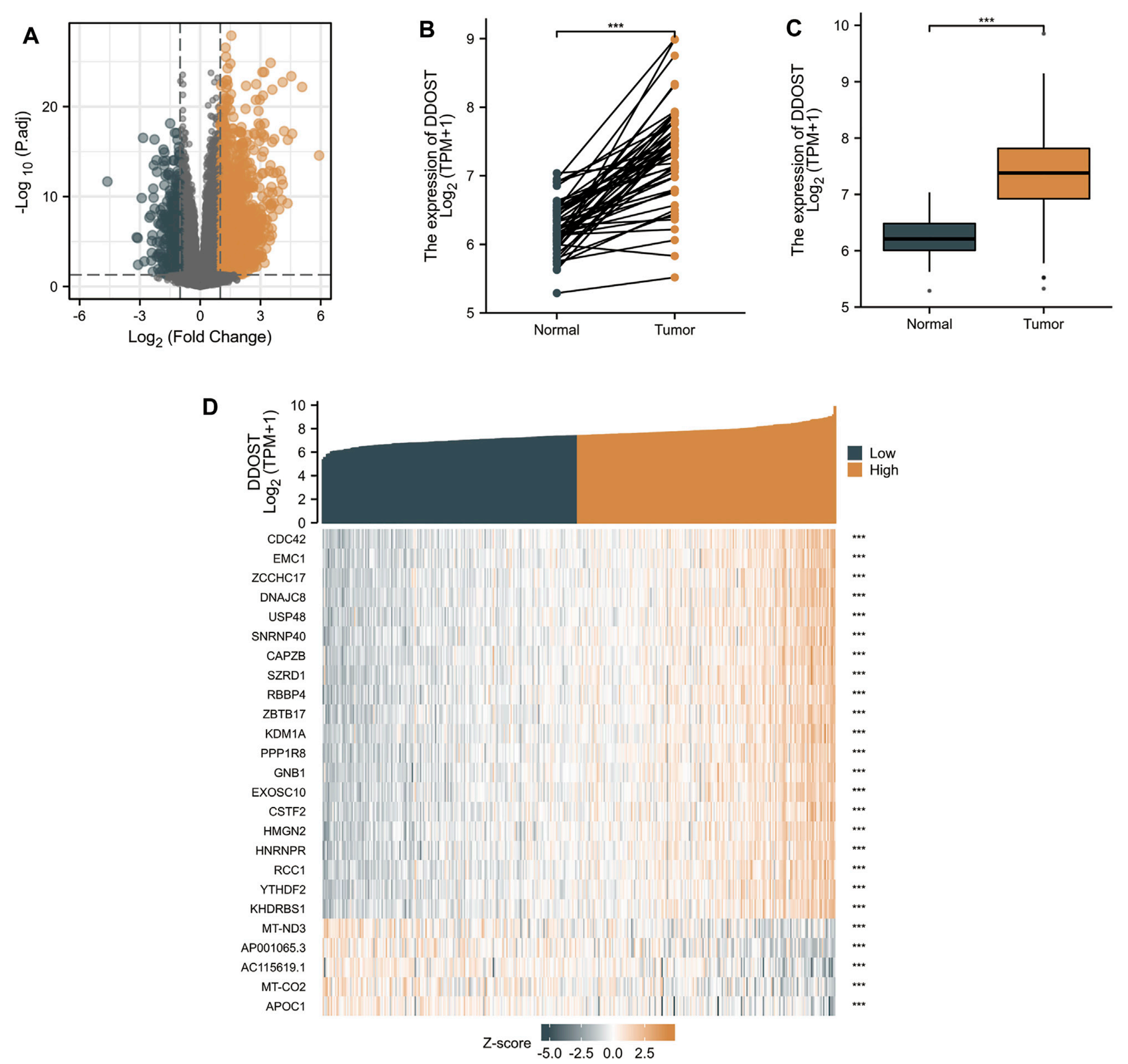


FIGURE 1 | The results of differentially expressed gene (DEG) analysis. **(A)** The volcano plot of differentially expressed RNAs. **(B,C)** The different expressions of DDOST between HCC and the normal group. **(D)** The heat map of the 25 genes correlated to DDOST.

RESULT

Clinical Characteristics

The clinical data of 374 HCC patients included the patients' age, gender, T stage, N stage, M stage, pathologic stage, gender, age, histologic grade, vascular invasion, OS event, BMI, and AFP (ng/ml) (**Table 1**). A total of 253 males and 121 females were analyzed in the present study. The Fisher's exact test result showed that DDOST was significantly correlated with OS event ($p = 0.013$); the chi-square test result revealed that DDOST had a trend of correlation with T stage ($p = 0.074$), histological grade ($p = 0.077$), and vascular invasion ($p = 0.072$). The Wilcoxon rank-sum test showed that DDOST was significantly correlated with AFP (ng/ml) ($p < 0.001$) and BMI ($p = 0.031$). DDOST expression was not significantly correlated with other clinicopathologic features.

Differential Expression Analysis of DDOST in HCC

With $|\text{logFC}| < 1.5$ and adjusted $p < 0.05$ set as the cut-off criteria, a total of 951 DEGs were identified (857 upregulated and 94 downregulated) by analyzing the HTSeq-Counts data of DDOST-related genes from TCGA. DEGs expressions were visualized in a

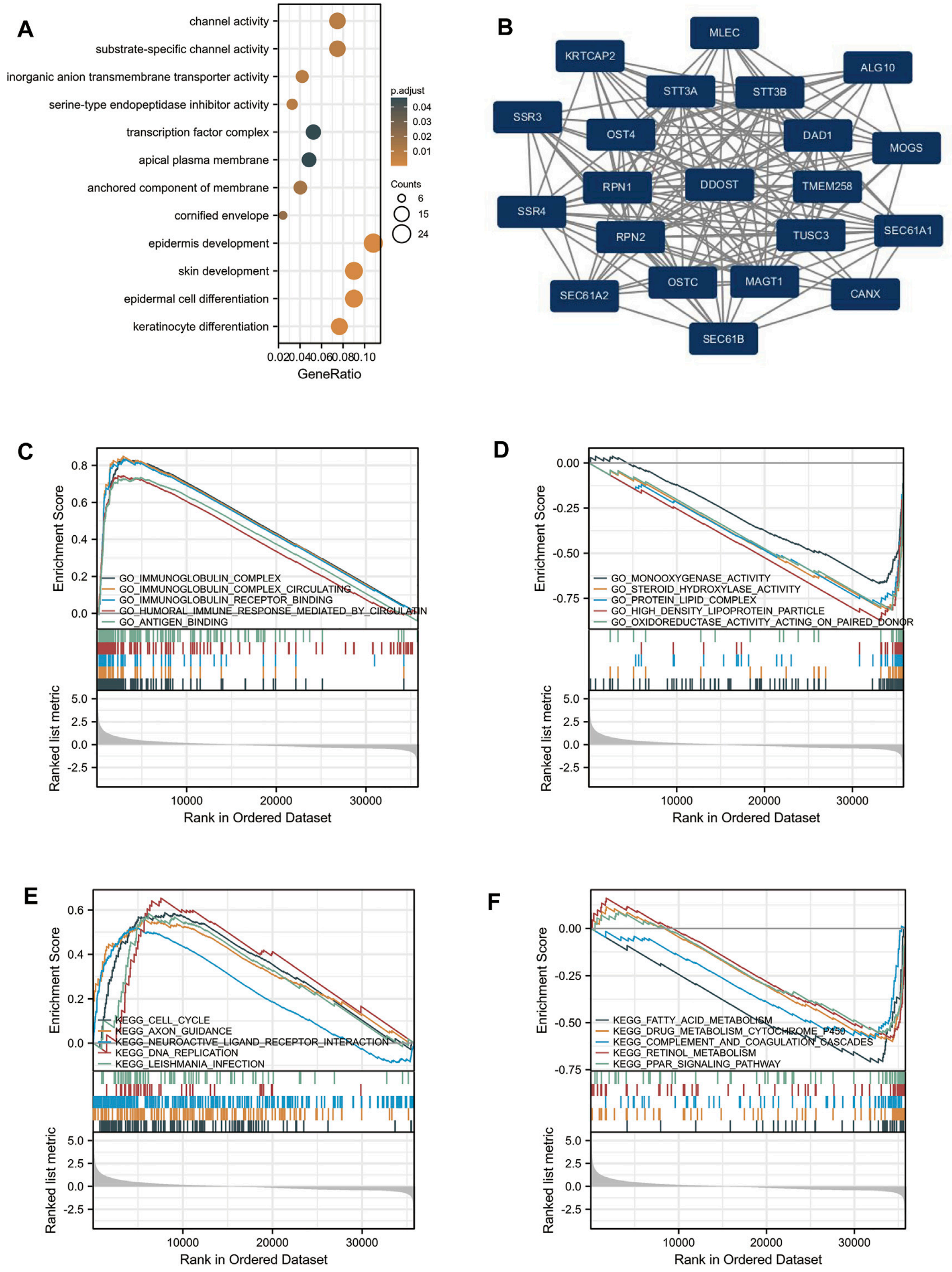


FIGURE 2 | Enrichment analysis of DDOST in HCC. **(A)** Biological process enrichment related to DDOST-related genes. **(B)** A network of DDOST and its 20 potential co-interaction proteins. **(C–F)** The results of enrichment analysis from GSEA.

volcano plot (**Figure 1A**). The unpaired and paired differential expression analyses between normal and HCC groups indicated that DDOST was expressed significantly higher in tumors compared to normal tissue (**Figures 1B,C**). The correlation between DDOST and 25 genes was demonstrated in a heat map (**Figure 1D**).

TABLE 2 | Signaling pathways most significantly associated with DDOST expression.

	Description	NES	P-value	p.adjust
Positive GO term	GO_IMMUNOGLOBULIN_COMPLEX	3.258	0.001	0.020
	GO_IMMUNOGLOBULIN_COMPLEX_CIRCULATING	2.969	0.001	0.020
	GO_IMMUNOGLOBULIN_RECEPTOR_BINDING	2.933	0.001	0.020
	GO_HUMORAL_IMMUNE_RESPONSE_MEDIATED_BY_CIRCULATING_IMMUNOGLOBULIN	2.889	0.001	0.020
	GO_ANTIGEN_BINDING	2.886	0.001	0.020
Negative GO term	GO_MONOOXYGENASE_ACTIVITY	−3.133	0.006	0.037
	GO_STEROID_HYDROXYLASE_ACTIVITY	−3.069	0.003	0.027
	GO_PROTEIN_LIPID_COMPLEX	−3.013	0.003	0.027
	GO_HIGH_DENSITY_LIPOPROTEIN_PARTICLE	−3.002	0.003	0.027
	GO_OXIDOREDUCTASE_ACTIVITY_ACTING_ON_PAIRED_DONORS	−2.988	0.003	0.027
Positive KEGG term	KEGG_CELL_CYCLE	2.273	0.001	0.025
	KEGG_AXON_GUIDANCE	2.173	0.001	0.025
	KEGG_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION	2.147	0.001	0.025
	KEGG_DNA_REPLICATION	2.084	0.003	0.026
	KEGG_LEISHMANIA_INFECTION	2.080	0.001	0.025
Negative KEGG term	KEGG_FATTY_ACID_METABOLISM	−2.817	0.004	0.027
	KEGG_DRUG_METABOLISM_CYTOCHROME_P450	−2.614	0.005	0.029
	KEGG_COMPLEMENT_AND_COAGULATION_CASCADES	−2.571	0.005	0.029
	KEGG_RETINOL_METABOLISM	−2.544	0.005	0.029
	KEGG_PPAR_SIGNALING_PATHWAY	−2.441	0.005	0.029

Functional Enrichment Analysis of DEGs

GO analysis indicated that DEG-related DDOST had significant regulation on epidermis development, skin development, epidermal cell differentiation, keratinocyte differentiation, channel activity, substrate-specific channel activity, inorganic anion transmembrane transporter activity, serine-type endopeptidase inhibitor activity, the anchored component of membrane, and cornified envelope (**Figure 2A**). The network of DDOST and its potential co-expression genes in DDOST-related DEGs are shown in **Figure 2B**.

To further identify the biological function of DDOST, the GSEA of differences between low and high DDOST expression data sets were performed to identify the GO term and KEGG pathway associated with DDOST. A total of 476 pathways showed significant differences (FDR <0.05, adjusted $p < 0.05$) in the enrichment of GO terms and KEGG pathways in samples with a high expression of DDOST. The most significantly enriched GO term and KEGG pathway based on their NES are shown in **Table 2**. The GSEA analysis in GO term revealed that immunoglobulin complex, immunoglobulin complex circulating, immunoglobulin complex receptor binding, humoral immune response mediated by circulating immunoglobulin, and antigen-binding were positively correlated with high levels of DDOST (**Figure 2C**); monooxygenase activity, steroid hydroxylase activity, protein–lipid complex, high-density lipoprotein particle, and oxidoreductase activity on paired donors were negatively correlated with high levels of DDOST (**Figure 2D**). The GSEA analysis in the KEGG pathway revealed that cell cycle, axon guidance, neuroactive ligand–receptor interaction, DNA replication, and *Leishmania* infection were positively correlated with high levels of DDOST (**Figure 2E**); Fatty acid metabolism, drug metabolism cytochrome P450, complement and coagulation cascades, retinol metabolism, and the PPAR signaling pathway were negatively correlated with high levels of DDOST (**Figure 2F**). These results indicate that the pathways regulating immunoglobulin complex, cell cycle control, and DNA replication were strongly associated with DDOST expression.

Relationship Between DDOST Expression and Immune Infiltration

Spearman correlation was employed to study the correlation between the DDOST expression level in TPM format and the immune cell infiltration level quantified as the ssGSEA score. The Th2 cells' infiltration level displays a significantly positive correlation with DDOST expression (Spearman R = 0.390, $p < 0.001$) (**Figure 3A**) and was significantly higher in the DDOST high-expression group ($p < 0.001$) (**Figure 3B**). On the other hand, the cytotoxic cells' infiltration level showed a significantly negative correlation with DDOST expression (Spearman R = -0.232, $p < 0.001$) (**Figure 3C**) and was significantly lower in the DDOST high-expression group ($p < 0.001$) (**Figure 3D**). T helper cells, NK CD56bright cells, TFH, aDC, and macrophages have also shown a positive relation with DDOST. pDC, DC, CD8 T cells, Th17 cells, Tgd, neutrophils, and NK cells have shown a negative correlation with DDOST (**Figure 3E**). These results indicated the vital role of DDOST in the immune infiltration in HCC. Different degrees of correlation between the ratios of 24 types of different tumor-infiltrating immune cells' subpopulations were assessed and visualized by a heat map (**Figure 3F**).

Associations Between DDOST Expression and Clinicopathologic Variables

Welch one-way ANOVA followed by the Bonferroni correction proved that the expression of DDOST was significantly correlated

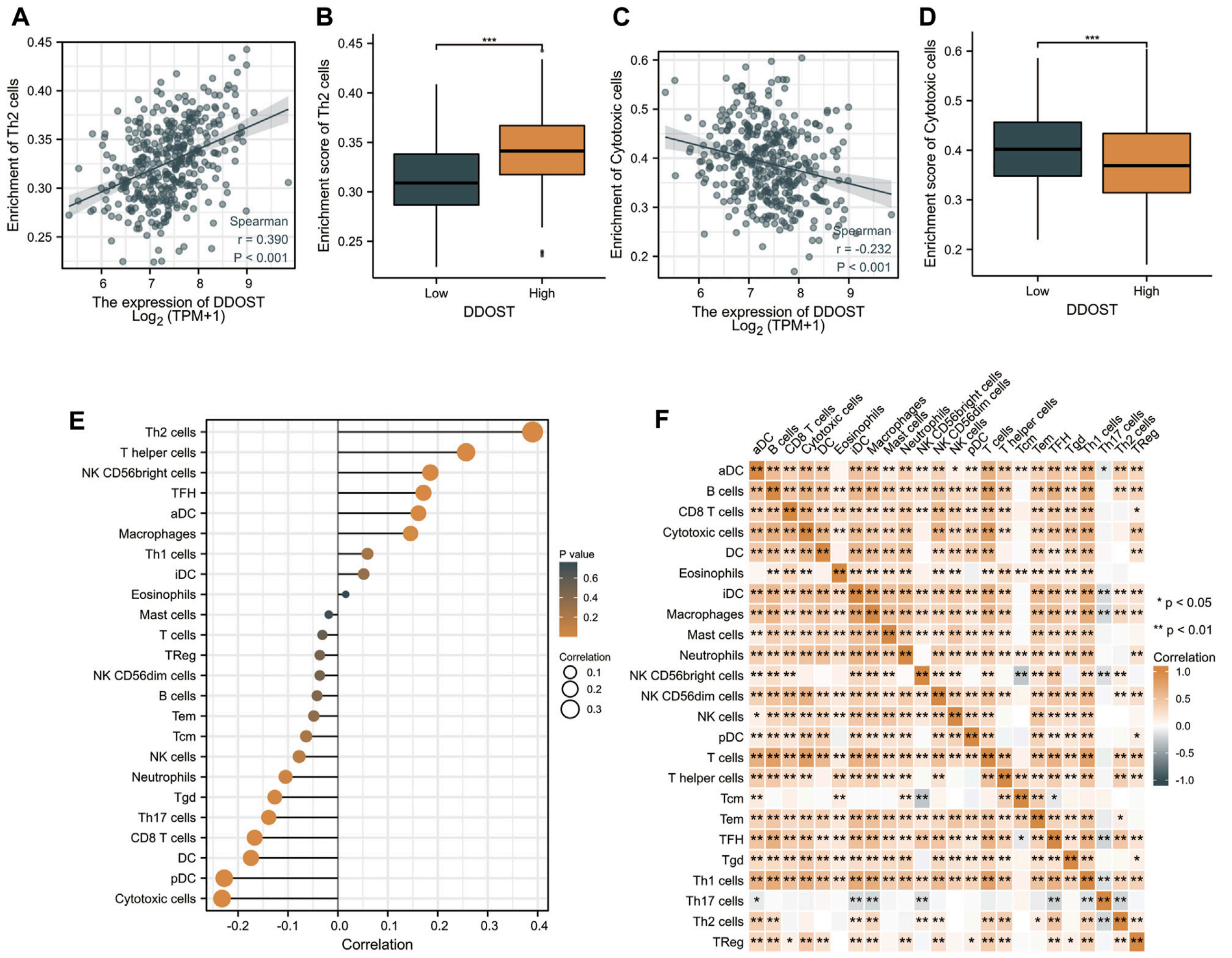


FIGURE 3 | The results of analysis between DDOST expression and immune infiltration. **(A)** The positive correlation between DDOST expression and Th2 cells. **(B)** Th2 cells' infiltration level in different DDOST expression groups. **(C)** The negative correlation between DDOST expression and cytotoxic cells. **(D)** Cytotoxic cells' infiltration level in different DDOST expression groups. **(E)** Correlation between DDOST expression level and the relative abundances of 24 immune cells. **(F)** Heat map of 24 immune infiltration cells in HCC.

with the pathologic stage and T stage (**Figures 4A,B**). The *t*-test revealed that the expression of DDOST was significantly correlated with the histologic grade, vascular invasion, and OS event (**Figures 4C–E**). Logistic regression analysis showed that DDOST was significantly correlated with the T stage ($p = 0.011$) and histologic grade ($p < 0.013$) and had a trend of correlation with vascular invasion ($p = 0.056$) (**Table 3**).

In the Cox regression model, univariate Cox regression indicates that the T stage ($p < 0.001$), M stage ($p = 0.017$), pathologic stage ($p < 0.001$), and DDOST ($p < 0.001$) were correlated with the bad prognosis of HCC (**Table 4**). All variables in univariate Cox regression were included in multivariate Cox regression. Multivariate Cox regression showed that T stage ($p = 0.017$) and DDOST ($p = 0.038$) were independent prognostic factors for OS (**Figure 5A**).

The distribution of DDOST expression, survival status of HCC patients, and expression profiles of DDOST are shown in **Figure 5B**. The blue dots represent the surviving HCC patients, and the orange dots represent the dead HCC patients. The upper line represents the median of risk score. The left side of the upper line represents the low-risk score group with a low expression of DDOST, and the right side of the dotted line represents the high-risk score group with a high expression of DDOST. With the increase of risk score in HCC patients, the number of orange dots increased gradually, and the number of dead HCC patients increased. It shows that the patients in the high-risk group have poorer survival and a higher risk of death.

The ROC analysis of DDOST supported the diagnostic accuracy of the score (AUC = 0.93, 95% CI: 0.903–0.956) (**Figure 5C**). The time-dependent accuracy of DDOST in predicting OS in 1, 2, and 3 years was also assessed through a time-dependent ROC analysis (**Figure 5D**).

The K-M survival curve drawn by survminer package in R was used to evaluate the prognostic value of DDOST in OS of HCC.

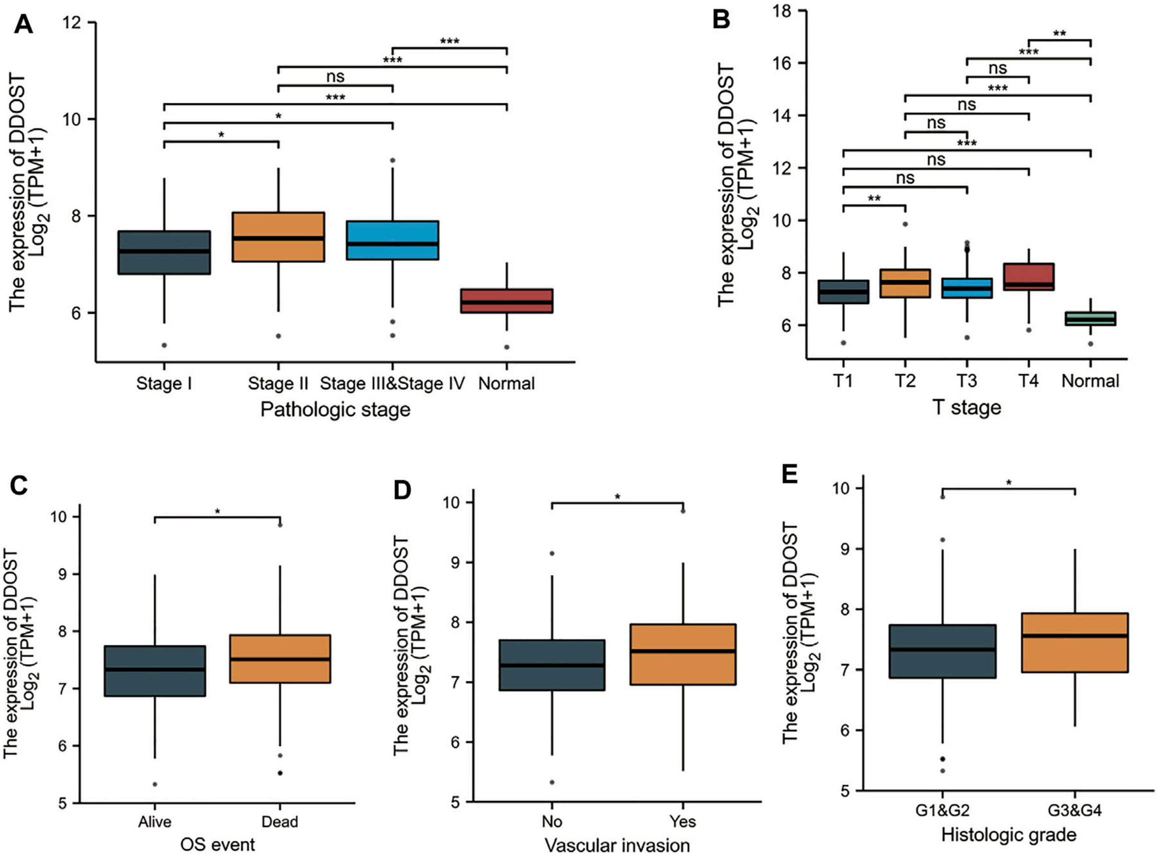


FIGURE 4 | Association between the DDOST expression and different clinicopathologic characteristics. **(A)** Association between the DDOST expression and the pathologic stage of HCC, **(B)** T stage, **(C)** OS event, **(D)** vascular invasion, and **(E)** histologic grade.

TABLE 3 | DDOST expression correlated with clinicopathological characteristics analyzed by logistic regression.

Characteristics	Total (N)	Odds ratio (OR)	*p*-Value
T stage (T2 and T3 and T4 vs. T1)	371	1.702 (1.130–2.572)	0.011
N stage (N1 vs. N0)	258	3.048 (0.384–62.060)	0.337
M stage (M1 vs. M0)	272	0.985 (0.117–8.308)	0.988
Pathologic stage (Stage III and Stage IV vs. Stage I and Stage II)	350	1.436 (0.888–2.333)	0.141
Histologic grade (G3 and G4 vs. G1 and G2)	369	1.719 (1.123–2.644)	0.013
Vascular invasion (Yes vs. No)	318	1.573 (0.990–2.511)	0.056
AFP (ng/ml) (>400 vs. ≤400)	280	1.598 (0.916–2.808)	0.100
Albumin (g/dl) (≥3.5 vs. <3.5)	300	0.751 (0.437–1.287)	0.297
Tumor status (with tumor vs. tumor-free)	355	1.301 (0.854–1.985)	0.221

HCC patients were divided into high and low expression groups based on the DDOST expression median value. The high expression group has a strong correlation with worse OS (HR = 1.96 (1.38–2.79), $p < 0.001$) and DSS (HR = 1.97 (1.25–3.09), $p = 0.003$) (**Figures 6A,B**).

The high expression of DDOST was also associated with worse OS in the T1 subgroup of T stage (HR = 2.06 (1.13–3.77), $p = 0.019$), stage I subgroup of pathologic stage (HR = 1.97 (1.05–3.68), $p = 0.035$), G1 and G2 subgroup of histologic grade [HR = 1.81 (1.14–2.86), $p = 0.011$], tumor- free subgroup of tumor status (HR = 2.14 (1.15–3.98), $p = 0.017$), and no vascular invasion subgroup of vascular invasion (HR = 1.97 (1.16–3.32), $p = 0.012$) (**Figures 6C–F**).

Data Validation

In all three GEO datasets, DDOST mRNA expression exhibited a significant increase in HCC when compared to the normal group (*p*-value < 0.01, Supplement **Figures 1A–C**). K-M survival plots also showed the group with high DDOST expression having poor OS rates (log rank *p*-value = 0.010, Supplement **Figure 1D**).

TABLE 4 | Univariate and multivariate analyses of clinical pathological parameters in HCC patients.

Characteristics	Total (N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	*p*-Value	Hazard ratio (95% CI)	*p*-Value
Age	373				
≤60	177	Reference			
>60	196	1.205 (0.850–1.708)	0.295	1.323 (0.811–2.159)	0.262
Gender	373				
Female	121	Reference			
Male	252	0.793 (0.557–1.130)	0.200	0.993 (0.597–1.652)	0.979
Histologic grade	368				
G1	55	Reference			
G2	178	1.162 (0.686–1.968)	0.577	0.810 (0.390–1.685)	0.573
G3 and G4	135	1.222 (0.710–2.103)	0.469	1.003 (0.489–2.058)	0.993
T stage	370				
T1 and T2	277	Reference			
T3 and T4	93	2.598 (1.826–3.697)	**<0.001**	2.183 (1.150–4.141)	**0.017**
M stage	272				
M0	268	Reference			
M1	4	4.077 (1.281–12.973)	**0.017**	2.152 (0.615–7.535)	0.231
N stage	258				
N0	254	Reference			
N1	4	2.029 (0.497–8.281)	0.324	1.561 (0.357–6.826)	0.554
DDOST	373	1.585 (1.241–2.026)	**<0.001**	1.491 (1.022–2.176)	**0.038**
Pathologic stage	349				
Stage I	173	Reference			
Stage II and Stage III and Stage IV	176	2.090 (1.429–3.055)	**<0.001**	1.490 (0.750–2.960)	0.255

Bold values were statistically significant

DISCUSSION

As far as we know, the majority of the membrane and secretory proteins synthesized in the ER are modified with N-glycans in eukaryotes. The N-glycosylation reaction catalyzed by OST had been implicated in cell-to-cell communication, signal transduction, trafficking, folding, and the degradation of proteins (Ohtsubo and Marth, 2006; Harada et al., 2015; Mikolajczyk et al., 2020) and was involved in the mechanism of tumor immune escape in the tumor microenvironment (Hsu et al., 2018). Thus, OST can be a potential therapeutic target for cancer treatment (Harada et al., 2019). RPN2, TUSC, as well as DDOST are the subunits of OST. The expression of RPN2 is positively correlated with the progression of breast cancers (Ono et al., 2015), non-small cell lung (Fujita et al., 2015), gastric (Fujimoto et al., 2018), esophageal (Li et al., 2019), and colorectal cancers (Bi and Jiang, 2018), whereas TUSC3 was reported as a candidate tumor suppressor (Vašíčková et al., 2018). DDOST, which also acted as advanced glycation end product-receptor 1 (Li et al., 1996), had been reported in regulating AGE, which increased oxidative stress and inflammation and may be involved in liver injury and subsequent carcinogenesis (Moy et al., 2013). In this research, we analyzed the sequencing data on liver cancer patients from TCGA to study the potential function and evaluate the prognostic value of DDOST.

DDOST is highly expressed in HCC patients and correlated with several advanced clinical features (pathological stage, T stage histologic grade, vascular invasion, OS event), which suggested that DDOST is a potential prognostic and diagnostic marker deserving further clinical validation. The function of DDOST in HCC was further investigated in GSEA using TCGA data.

The PPI network indicated that DDOST can interact with SSR3, SSR4, SEC61A1, SEC61A2, SEC61B, CANX, MOGS, ALG10, and MLEC other than the subunit protein of OST. All those proteins were closely associated with the N-linked oligosaccharide processing pathway, which had long been considered directly associated with the metastatic potential of malignant tumor cells (Dennis, 1991). GSEA showed that positively enriched GO terms including immunoglobulin complex, immunoglobulin complex circulating, immunoglobulin complex receptor binding, humoral immune response mediated by circulating immunoglobulin, and antigen-binding were pivotal in immune complex formation. It is indicated that DDOST might participate in the immune response in the tumorigenesis of HCCs. On the other hand, the KEGG pathway analysis indicated that cancer-related pathways including cell cycle and DNA replication were positively enriched when DDOST was highly expressed. Furthermore, GSEA analysis revealed the vital role of DDOST in the metabolism of protein and fatty acids, and in the downregulation of the PPAR signaling pathway, which plays a vital role in protecting the liver from oxidation, inflammation, fibrosis, and tumors (Wu et al., 2020). Based on the results, we can presume that as a crucial molecular in regulating protein and lipid metabolism, the high DDOST expression may induce an immune response and regulate cell cycles in HCC by suppressing the PPAR pathway.

ssGSEA combined with Spearman correlation was adopted to investigate the relationship between DDOST expression and immune infiltration levels in HCC. Our results demonstrate that DDOST expression has a significantly positive correlation

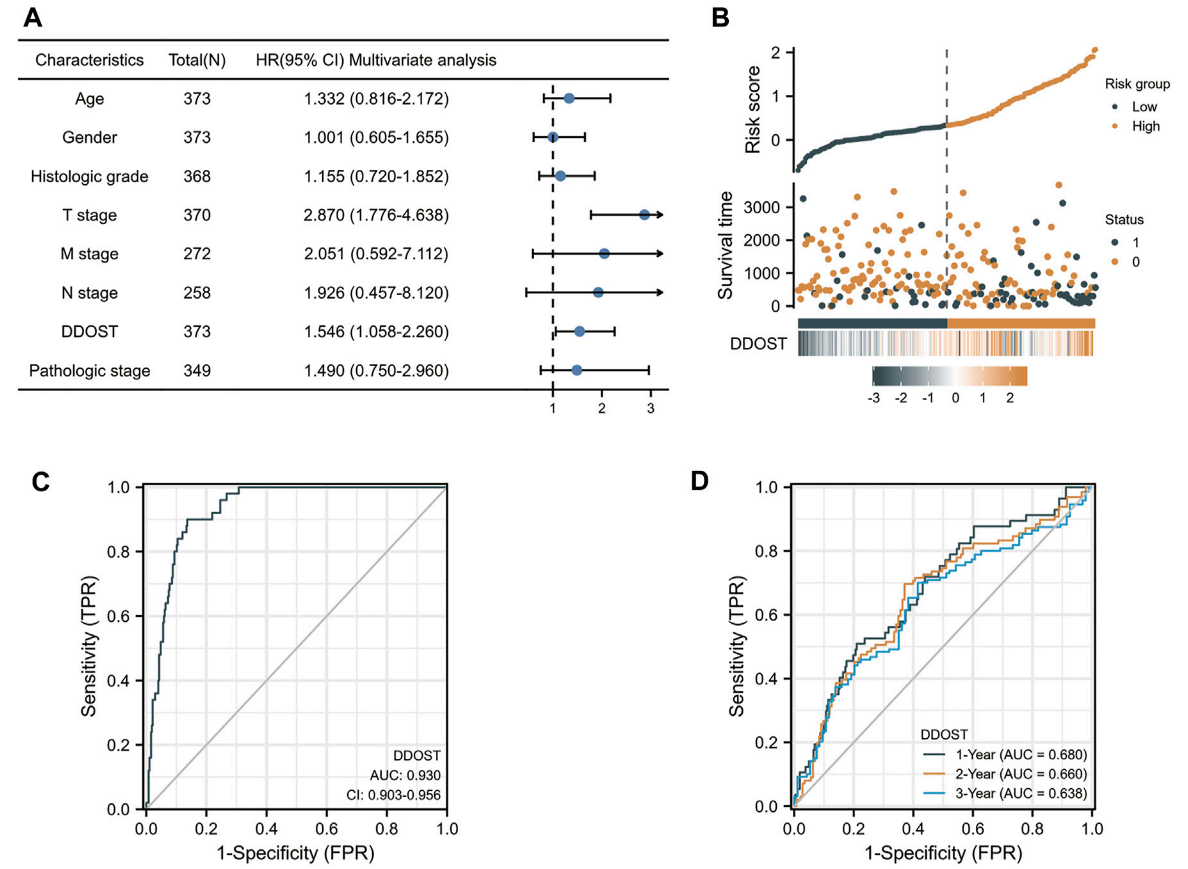


FIGURE 5 | The prognostic value of DDOST in LIHC. **(A)** Multivariate Cox regression visualized in the forest plot **(B)** DDOST expression distribution and survival status. 0: dead, 1: alive. **(C)** Diagnostic ROC curve of DDOST. **(D)** Time-dependent ROC curve of DDOST.

with Th2 cells and a strong-to-moderate correlation with T helper cells, NK CD56 bright cells, Tfh, aDC, macrophages, and Th1 cells. Our results indicate that a shift of Th1/Th2 balance toward Th2, which plays a vital role in HCC metastasis (Budhu and Wang, 2006), may be caused by the DDOST high expression. Th2 cell is one type of T helper cell that can induce the polarization of M1 macrophages into immunosuppressive M2 macrophages (DeNardo et al., 2009), and lead to the inhibition of the host immune system, hence contributing to tumorigenesis. IL-4 produced by Th2 cells can result in the activation of several cancer-related pathways (Zhao et al., 2015; Dey et al., 2020). Tfh cells can differentiate into Th1 and Th2 cells and regulate humoral immune response (Rezende et al., 2018). Based on previous research, we can conclude that overexpression of DDOST may induce immune infiltration in HCC genesis and progression.

There is an inverse correlation between cytotoxic cells, pDC, DC, CD8 T cells, Th17 cells, Tgd, neutrophils, NK cells, and DDOST. Cytotoxic cells including NK cells play a vital role in anti-tumor immunity. NK cells are important in innate immune surveillance against cancer (Lanier, 2005). CD8$^+$ T cells exhibit a cytotoxic ability against tumor cells through differentiating cytotoxic T cells (Iwahori, 2020). DCs including pDC were essential contributors to immune defenses against cancer. IFN-I produced by pDC shows good anti-tumor activity (Saulep-Easton et al., 2014). Th17 cells were closely related to neutrophils (Amicarella et al., 2017), and they are critical in tumor immunity and predict a poor prognosis in HCC (Wang et al., 2020). The downregulation of those types of immune cells may facilitate the progression of HCC. All findings according to ssGSEA exhibited the important role of DDOST in regulating immune infiltration in HCC.

As a traditional serological marker, AFP has been adopted in the diagnosis of HCC for decades (Wong et al., 2014). However, AFP was not significant in all HCC cases. It is reported that only 60%–70% of total HCC patients have elevated AFP levels, and nonspecific increases are also observed in non-HCC diseases such as chronic hepatitis or liver cirrhosis (Akeyama et al., 1972; Di Bisceglie and Hoofnagle, 1989). More importantly, AFP levels are usually normal in early HCC (Chen et al., 1984). On the contrary, DDOST is highly expressed in early-stage HCC. Furthermore, compared with the DDOST low-level expression group, HCC patients with DDOST highly expressed the result in poor OS and DSS. The Cox HR model also suggested that DDOST was strongly associated with the OS in HCC. The relation between DDOST and those prognostic indicators suggested that DDOST was a powerful prognostic biomarker in HCC.

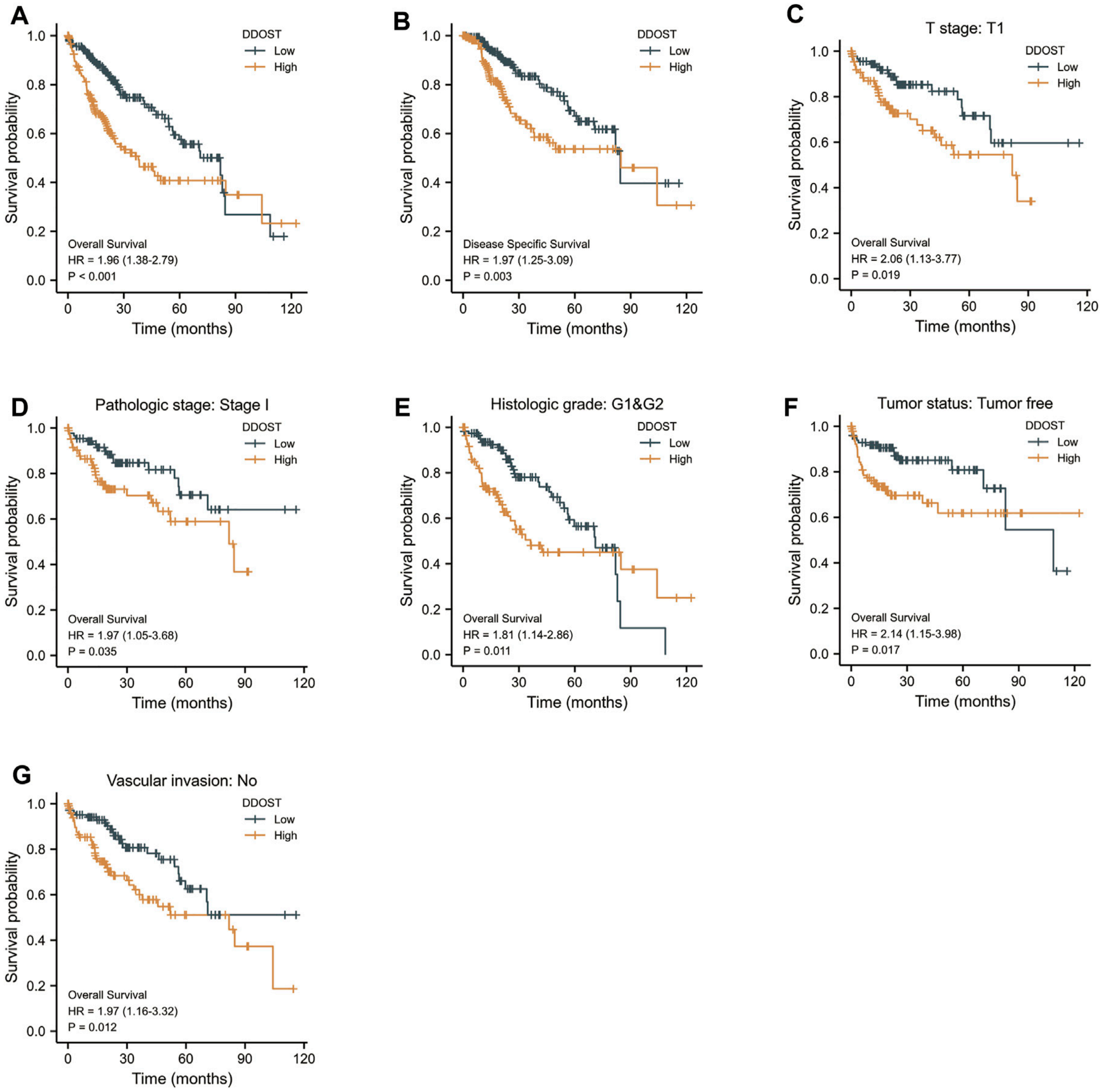


FIGURE 6 | The prognostic value of DDOST in the different subgroups. **(A,B)** The prognostic value of DDOST in OS and DSS of HCC. **(C–G)** High expression of DDOST was associated with worse OS in different subgroups.

Overall, the important role of DDOST in HCC was revealed through our study. Our work demonstrated that increased expression of DDOST is associated with poor OS in HCC patients. GSEA showed that pathways including DNA replication, cell cycle, immune response in cancer, the PPAR signaling pathway, and lipid acid metabolism were associated with DDOST expression. Moreover, the connection between DDOST and tumor-infiltrating immune cells was identified. The work presented here provides a detailed analysis of the role of DDOST in HCC development, which will aid in the understanding of the mechanisms underlying HCC. Combined with a previous study, our research indicates that DDOST can affect protein and lipid metabolism by joining in nascent polypeptide processing in ER and regulating the AGE level; then, it may lead to PPAR pathway suppression and play an important role in cell cycle regulation and immune infiltration in HCC.

Although the vital role of DDOST in the regulation of the cell cycle and immune response in the tumorigenesis of HCC had been proven, *in vitro* and *in vivo* experiments are still needed to verify the correlation between DDOST expression and HCC development and then to illustrate the biological mechanism of DDOST in HCC progression. Clinical researches are required to evaluate the relationship between DDOST expression and clinical features including the HCC stage, and prognosis value, which might facilitate the identification of new markers for assessing the tumor progression, promoting drug development, and improving treatment strategy.

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

CZ and RT conceived and designed the study. CZ, XJ, and HX did the literature review and conceptualization. CZ and JG analyzed the data and wrote the original draft. All authors read and approved the final manuscript.

FUNDING

We are grateful for support of Key Research and Development Program of Science and Technology Department of Sichuan Province (2019YFS0514) and National Key R and D Program of China (2020YFC2005500).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.819520/full#supplementary-material

REFERENCES

Akeyama, T., Koyama, T., and Kamada, T. (1972). Alpha-fetoprotein in Acute Viral Hepatitis. *N. Engl. J. Med.* 287, 989. doi:10.1056/nejm197211092871923

Amicarella, F., Muraro, M. G., Hirt, C., Cremonesi, E., Padovan, E., Mele, V., et al. (2017). Dual Role of Tumour-Infiltrating T Helper 17 Cells in Human Colorectal Cancer. *Gut* 66, 692–704. doi:10.1136/gutjnl-2015-310016

Bi, C., and Jiang, B. (2018). Downregulation of RPN2 Induces Apoptosis and Inhibits Migration and Invasion in colon Carcinoma. *Oncol. Rep.* 40, 283–293. doi:10.3892/or.2018.6434

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity* 39, 782–795. doi:10.1016/j.immuni.2013.10.003

Blum, A., Wang, P., and Zenklusen, J. C. (2018). SnapShot: TCGA-Analyzed Tumors. *Cell* 173, 530. doi:10.1016/j.cell.2018.03.059

Budhu, A., and Wang, X. (2006). The Role of Cytokines in Hepatocellular Carcinoma. *J. Leukoc. Biol.* 80, 1197–1213. doi:10.1189/jlb.0506297

Chen, D.-S., Sung, J.-L., Shed, J.-C., Lai, M.-Y., How, S.-W., Hsu, H.-C., et al. (1984). Serum α-Fetoprotein in the Early Stage of Human Hepatocellular Carcinoma. *Gastroenterology* 86, 1404–1409. doi:10.1016/s0016-5085(84)80151-1

DeNardo, D. G., Barreto, J. B., Andreu, P., Vasquez, L., Tawfik, D., Kolhatkar, N., et al. (2009). CD4+ T Cells Regulate Pulmonary Metastasis of Mammary Carcinomas by Enhancing Protumor Properties of Macrophages. *Cancer cell* 16, 91–102. doi:10.1016/j.ccr.2009.06.018

Dennis, J. W. (1991). N-linked Oligosaccharide Processing and Tumor Cell Biology. *Semin. Cancer Biol.* 2, 411–420.

Dey, P., Li, J., Zhang, J., Chaurasiya, S., Strom, A., Wang, H., et al. (2020). Oncogenic KRAS-Driven Metabolic Reprogramming in Pancreatic Cancer Cells Utilizes Cytokines from the Tumor Microenvironment. *Cancer Discov.* 10, 608–625. doi:10.1158/2159-8290.CD-19-0297

Di Bisceglie, A. M., and Hoofnagle, J. H. (1989). Elevations in Serum Alpha-Fetoprotein Levels in Patients with Chronic Hepatitis B. *Cancer* 64, 2117–2120. doi:10.1002/1097-0142(19891115)64:10<2117:aid-cncr2820641024>3.0.co;2-7

Forner, A., Reig, M., and Bruix, J. (2018). Hepatocellular Carcinoma. *Lancet* 391, 1301–1314. doi:10.1016/S0140-6736(18)30010-2

Fu, J., and Kreibich, G. (2000). Retention of Subunits of the Oligosaccharyltransferase Complex in the Endoplasmic Reticulum. *J. Biol. Chem.* 275, 3984–3990. doi:10.1074/jbc.275.6.3984

Fujimoto, D., Goi, T., Koneri, K., and Hirono, Y. (2018). RPN2 Is Effective Biomarker to Predict the Outcome of Combined Chemotherapy Docetaxel and Cisplatin for Advanced Gastric Cancer. *Oncotarget* 9, 15208–15218. doi:10.18632/oncotarget.24622

Fujita, Y., Yagishita, S., Takeshita, F., Yamamoto, Y., Kuwano, K., and Ochiya, T. (2015). Prognostic and Therapeutic Impact of RPN2-Mediated Tumor Malignancy in Non-small-cell Lung Cancer. *Oncotarget* 6, 3335–3345. doi:10.18632/oncotarget.2793

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7

Harada, Y., Hirayama, H., and Suzuki, T. (2015). Generation and Degradation of Free Asparagine-Linked Glycans. *Cell. Mol. Life Sci.* 72, 2509–2533. doi:10.1007/s00018-015-1881-7

Harada, Y., Ohkawa, Y., Kizuka, Y., and Taniguchi, N. (2019). Oligosaccharyltransferase: A Gatekeeper of Health and Tumor Progression. *Int. J. Mol. Sci.* 20, 6074. doi:10.3390/ijms20236074

Hlady, R. A., Sathyanarayan, A., Thompson, J. J., Zhou, D., Wu, Q., Pham, K., et al. (2019). Integrating the Epigenome to Identify Drivers of Hepatocellular Carcinoma. *Hepatology* 69, 639–652. doi:10.1002/hep.30211

Hollenbach, M. (2017). The Role of Glyoxalase-I (Glo-I), Advanced Glycation Endproducts (AGEs), and Their Receptor (RAGE) in Chronic Liver Disease and Hepatocellular Carcinoma (HCC). *Int. J. Mol. Sci.* 18, 2466. doi:10.3390/ijms18112466

Howell, J., Pedrana, A., Schroeder, S. E., Scott, N., Aufegger, L., Atun, R., et al. (2021). A Global Investment Framework for the Elimination of Hepatitis B. *J. Hepatol.* 74, 535–549. doi:10.1016/j.jhep.2020.09.013

Hsu, J.-M., Li, C.-W., Lai, Y.-J., and Hung, M.-C. (2018). Posttranslational Modifications of PD-L1 and Their Applications in Cancer Therapy. *Cancer Res.* 78, 6349–6353. doi:10.1158/0008-5472.CAN-18-1892

Iwahori, K. (2020). Cytotoxic CD8+ Lymphocytes in the Tumor Microenvironment. *Adv. Exp. Med. Biol.* 1224, 53–62. doi:10.1007/978-3-030-35723-8_4

Jones, M. A., Ng, B. G., Bhide, S., Chin, E., Rhodenizer, D., He, P., et al. (2012). DDOST Mutations Identified by Whole-Exome Sequencing Are Implicated in Congenital Disorders of Glycosylation. *Am. J. Hum. Genet.* 90, 363–368. doi:10.1016/j.ajhg.2011.12.024

Kelley, R. K., Meyer, T., Rimassa, L., Merle, P., Park, J.-W., Yau, T., et al. (2020). Serum Alpha-Fetoprotein Levels and Clinical Outcomes in the Phase III CELESTIAL Study of Cabozantinib versus Placebo in Patients with Advanced Hepatocellular Carcinoma. *Clin. Cancer Res.* 26, 4795–4804. doi:10.1158/1078-0432.CCR-19-3884

Lanier, L. L. (2005). NK Cell Recognition. *Annu. Rev. Immunol.* 23, 225–274. doi:10.1146/annurev.immunol.23.021704.115526

Li, Y., Huang, C., Bai, Q., and Yu, J. (2019). Ribophorin II Promotes Cell Proliferation, Migration, and Invasion in Esophageal Cancer Cells *In Vitro* and *In Vivo*. *Biosci. Rep.* 39, BSR20182448. doi:10.1042/BSR20182448

Li, Y. M., Mitsuhashi, T., Wojciechowicz, D., Shimizu, N., Li, J., Stitt, A., et al. (1996). Molecular Identity and Cellular Distribution of Advanced Glycation Endproduct Receptors: Relationship of P60 to OST-48 and P90 to 80K-H Membrane Proteins. *Proc. Natl. Acad. Sci.* 93, 11047–11052. doi:10.1073/pnas.93.20.11047

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Menyhárt, O., Nagy, Á., and Győrffy, B. (2018). Determining Consistent Prognostic Biomarkers of Overall Survival and Vascular Invasion in Hepatocellular Carcinoma. *R. Soc. Open Sci.* 5, 181006. doi:10.1098/rsos.181006

Mikolajczyk, K., Kaczmarek, R., and Czerwinski, M. (2020). How Glycosylation Affects Glycosylation: the Role of N-Glycans in Glycosyltransferase Activity. *Glycobiology* 30, 941–969. doi:10.1093/glycob/cwaa041

Moy, K. A., Jiao, L., Freedman, N. D., Weinstein, S. J., Sinha, R., Virtamo, J., et al. (2013). Soluble Receptor for Advanced Glycation End Products and Risk of Liver Cancer. *Hepatology* 57, 2338–2345. doi:10.1002/hep.26264

Ohtsubo, K., and Marth, J. D. (2006). Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* 126, 855–867. doi:10.1016/j.cell.2006.08.019

Ono, M., Tsuda, H., Kobayashi, T., Takeshita, F., Takahashi, R.-U., Tamura, K., et al. (2015). The Expression and Clinical Significance of Ribophorin II (RPN2) in Human Breast Cancer. *Pathol. Int.* 65, 301–308. doi:10.1111/pin.12297

Petrick, J. L., Florio, A. A., Znaor, A., Ruggieri, D., Laversanne, M., Alvarez, C. S., et al. (2020). International Trends in Hepatocellular Carcinoma Incidence, 1978-2012. *Int. J. Cancer* 147, 317–330. doi:10.1002/ijc.32723

Rezende, R. M., Lanser, A. J., Rubino, S., Kuhn, C., Skillin, N., Moreira, T. G., et al. (2018). Γδ T Cells Control Humoral Immune Response by Inducing T Follicular Helper Cell Differentiation. *Nat. Commun.* 9, 3151. doi:10.1038/s41467-018-05487-9

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., et al. (2011). pROC: an Open-Source Package for R and S+ to Analyze and Compare ROC Curves. *BMC Bioinformatics* 12, 77. doi:10.1186/1471-2105-12-77

Saulep-Easton, D., Vincent, F. B., Le Page, M., Wei, A., Ting, S. B., Croce, C. M., et al. (2014). Cytokine-driven Loss of Plasmacytoid Dendritic Cell Function in Chronic Lymphocytic Leukemia. *Leukemia* 28, 2005–2015. doi:10.1038/leu.2014.105

Shapanis, A., Lai, C., Smith, S., Coltart, G., Sommerlad, M., Schofield, J., et al. (2021). Identification of Proteins Associated with Development of Metastasis from Cutaneous Squamous Cell Carcinomas (cSCCs) via Proteomic Analysis of Primary cSCCs*. *Br. J. Dermatol.* 184, 709–721. doi:10.1111/bjd.19485

Singal, A. G., Lampertico, P., and Nahon, P. (2020). Epidemiology and Surveillance for Hepatocellular Carcinoma: New Trends. *J. Hepatol.* 72, 250–261. doi:10.1016/j.jhep.2019.08.025

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. doi:10.1073/pnas.0506580102

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-wide Experimental Datasets. *Nucleic Acids Res.* 47, D607–D613. doi:10.1093/nar/gky1131

Vašíčková, K., Horak, P., and Vaňhara, P. (2018). TUSC3: Functional Duality of a Cancer Gene. *Cel. Mol. Life Sci.* 75, 849–857. doi:10.1007/s00018-017-2660-4

Wang, Y., Yao, R., Zhang, D., Chen, R., Ren, Z., and Zhang, L. (2020). Circulating Neutrophils Predict Poor Survival for HCC and Promote HCC Progression through P53 and STAT3 Signaling Pathway. *J. Cancer* 11, 3736–3744. doi:10.7150/jca.42953

Wong, G. L. H., Chan, H. L. Y., Tse, Y.-K., Chan, H.-Y., Tse, C.-H., Lo, A. O. S., et al. (2014). On-treatment Alpha-Fetoprotein Is a Specific Tumor Marker for Hepatocellular Carcinoma in Patients with Chronic Hepatitis B Receiving Entecavir. *Hepatology* 59, 986–995. doi:10.1002/hep.26739

Wu, L., Guo, C., and Wu, J. (2020). Therapeutic Potential of PPARγ Natural Agonists in Liver Diseases. *J. Cel Mol. Med.* 24, 2736–2748. doi:10.1111/jcmm.15028

Yamagata, T., Tsuru, T., Momoi, M. Y., Suwa, K., Nozaki, Y., Mukasa, T., et al. (1997). Genome Organization of Human 48-kDa Oligosaccharyltransferase (DDOST). *Genomics* 45, 535–540. doi:10.1006/geno.1997.4966

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

Zhao, P., Bu, X., Wei, X., Sun, W., Xie, X., Li, C., et al. (2015). Dendritic Cell Immunotherapy Combined with Cytokine-Induced Killer Cells Promotes Skewing toward Th2 Cytokine Profile in Patients with Metastatic Non-small Cell Lung Cancer. *Int. Immunopharmacol.* 25, 450–456. doi:10.1016/j.intimp.2015.02.010

Zhao, X., Liu, J., Ge, S., Chen, C., Li, S., Wu, X., et al. (2019). Saikosaponin A Inhibits Breast Cancer by Regulating Th1/Th2 Balance. *Front. Pharmacol.* 10, 624. doi:10.3389/fphar.2019.00624

Zhuang, A., Yap, F. Y., Bruce, C., Leung, C., Plan, M. R., Sullivan, M. A., et al. (2017). Increased Liver AGEs Induce Hepatic Injury Mediated through an OST48 Pathway. *Sci. Rep.* 7, 12292. doi:10.1038/s41598-017-12548-4

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 16 February 2022
doi: 10.3389/fgene.2021.806386

Predicting Mutational Status of Driver and Suppressor Genes Directly from Histopathology With Deep Learning: A Systematic Study Across 23 Solid Tumor Types

Chiara Maria Lavinia Loeffler[1,2]*, Nadine T. Gaisa[3,2], Hannah Sophie Muti[1,2], Marko van Treeck[1,2], Amelie Echle[1,2], Narmin Ghaffari Laleh[1,2], Christian Trautwein[1,2], Lara R. Heij[3,4,5,2], Heike I. Grabsch[6,7], Nadina Ortiz Bruechle[3,2†] and Jakob Nikolas Kather[1,7,8,2†]

[1]Department of Medicine III, University Hospital RWTH Aachen, Aachen, Germany, [2]Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf (CIO ABCD), Aachen, Germany, [3]Institute of Pathology, University Hospital RWTH Aachen, Aachen, Germany, [4]Department of Surgery and Transplantation, University Hospital RWTH Aachen, Aachen, Germany, [5]NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht, Netherlands, [6]Department of Pathology, GROW School for Oncology and Reproduction, Maastricht University Medical Center+, Maastricht, Netherlands, [7]Pathology and Data Analytics, Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, United Kingdom, [8]Medical Oncology, National Center for Tumor Diseases (NCT), University Hospital Heidelberg, Heidelberg, Germany

Edited by:
Luis Zapata,
Institute of Cancer Research (ICR),
United Kingdom

Reviewed by:
Martin Schaefer,
European Institute of Oncology (IEO),
Italy
Francisco Martinez-Jimenez,
University Medical Center Utrecht,
Netherlands

***Correspondence:**
Chiara Maria Lavinia Loeffler
chiara.loeffler@rwth-aachen.de

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

***Received:** 31 October 2021*
***Accepted:** 30 December 2021*
***Published:** 16 February 2022*

Citation:
Loeffler CML, Gaisa NT, Muti HS, van Treeck M, Echle A, Ghaffari Laleh N, Trautwein C, Heij LR, Grabsch HI, Ortiz Bruechle N and Kather JN (2022) Predicting Mutational Status of Driver and Suppressor Genes Directly from Histopathology With Deep Learning: A Systematic Study Across 23 Solid Tumor Types. Front. Genet. 12:806386. doi: 10.3389/fgene.2021.806386

In the last four years, advances in Deep Learning technology have enabled the inference of selected mutational alterations directly from routine histopathology slides. In particular, recent studies have shown that genetic changes in clinically relevant driver genes are reflected in the histological phenotype of solid tumors and can be inferred by analysing routine Haematoxylin and Eosin (H&E) stained tissue sections with Deep Learning. However, these studies mostly focused on selected individual genes in selected tumor types. In addition, genetic changes in solid tumors primarily act by changing signaling pathways that regulate cell behaviour. In this study, we hypothesized that Deep Learning networks can be trained to directly predict alterations of genes and pathways across a spectrum of solid tumors. We manually outlined tumor tissue in H&E-stained tissue sections from 7,829 patients with 23 different tumor types from The Cancer Genome Atlas. We then trained convolutional neural networks in an end-to-end way to detect alterations in the most clinically relevant pathways or genes, directly from histology images. Using this automatic approach, we found that alterations in 12 out of 14 clinically relevant pathways and numerous single gene alterations appear to be detectable in tissue sections, many of which have not been reported before. Interestingly, we show that the prediction performance for single gene alterations is better than that for pathway alterations. Collectively, these data demonstrate the predictability of genetic alterations directly from routine cancer histology images and show that individual genes leave a stronger morphological signature than genetic pathways.

Keywords: deep learning, artificail intelligence (AI), cancer pathway, cancer pathway genes, genetic, TCGA

INTRODUCTION

Genetic changes can influence the cell and tissue morphology of solid tumors (**Figure 1A**). This morphology can be observed in routine histopathology images which are available for almost every patient with any solid tumor. Routinely, histopathologists review H&E stained tissue sections to establish a diagnosis, stage a disease etc. Due to recent advances in computer vision, automatic image analysis can extract subtle features from digital tissue sections which seem to be elusive to the human eye (Echle et al., 2020b). In particular, Deep Learning (DL), an artificial intelligence method, has been used to analyze histology images (Kather and Calderaro, 2020) and multiple studies demonstrated that Deep Learning can link morphological changes in cancer histology images to specific genetic alterations. Early studies in the field predicted clinically relevant genetic mutations in lung cancer (Coudray et al., 2018), colorectal cancer (Kather et al., 2019), breast cancer (Naik et al., 2020) and other tumor types from histological whole slide images. More recently, multiple studies suggested that many genetic alterations are predictable from routine histology alone across different tumor types (Fu et al., 2020; Kather et al., 2020; Schmauch et al., 2020; Loeffler et al., 2021; Muti et al., 2021). Previous studies focused on predicting single gene alterations. However, it is well known that certain gene products act together in functional pathways and mutations (MUT) of different genes of the same pathway may have a similar effect such as pathway activation (Ben-Hamo et al., 2020). To understand the effect of genetic alterations on tumor biology, potential genetic alterations need to be considered in the context of their functional significance in the affected

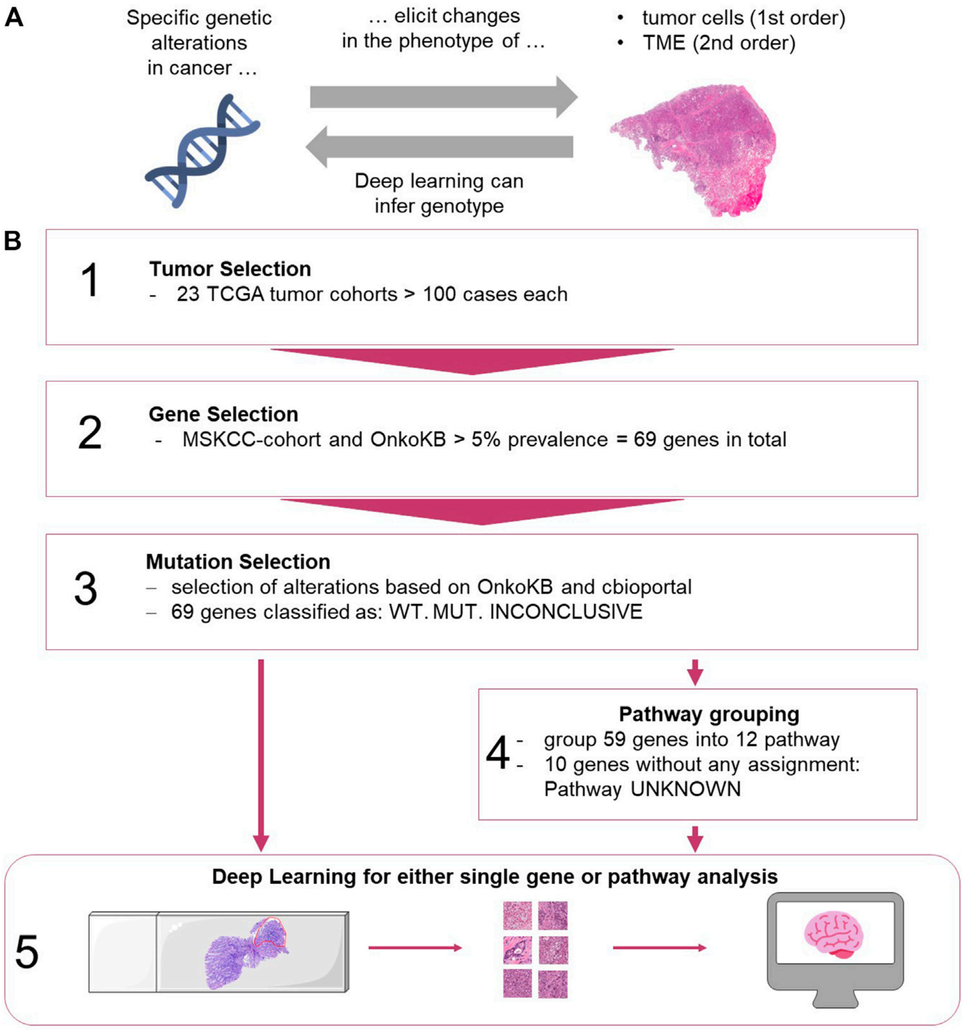


FIGURE 1 | (A) Biological hypothesis of this study. TME: tumor microenvironment. **(B)** Workflow for selection of data and Deep Learning methods. **(1)** Tumors from TCGA were analyzed. **(2)** Genes were selected based on the MSKCC cohort and OnkoKB platforms. **(3)** Alterations were grouped based on different sources. **(4)** Genes were grouped into pathways (see **Supplementary Table S1**). **(5)** Processing of images and training of the network for genes alone and grouped into pathways. (images from https://smart.servier.com, and Twitter Twemoji under a CC-BY license). TME: tumor microenvironment, TCGA: The Cancer Genome Atlas, MSKCC, OnkoKB, WT: wild type, MUT: mutation present.

pathway. For example, it has been shown that both, *PTEN* loss and *PIK3CA* mutation can lead to the activation of the *PI3K* or *MAPK* pathway in cancer of the breast, colorectum, stomach or lung (Dhillon et al., 2007; Jiang et al., 2020). This phenomenon, can be of therapeutic relevance, as targeted therapies may not only affect one specific gene, but also affect other downstream genes. Instead of focusing on a single gene, in some cases it might even be sufficient to identify pathway activation or inhibition to predict treatment response or failure (Schumacher et al., 2019; Ben-Hamo et al., 2020).

We hypothesized that alteration of a particular signaling pathway leads to histomorphological changes which can be predicted from routinely stained pathology slides using Deep Learning technology.

The aim of the current study was to systematically compare the predictability of an "overall altered signaling pathway" to a "single altered gene" of the same pathway. To this end, we analyzed the 69 most frequently mutated genes in 23 cancer types, representing 12 oncogenic pathways, and trained end-to-end Deep Learning networks to predict single gene mutations or signaling pathway alterations. Furthermore, we aimed to extend the evaluation of Deep Learning-based detection of genetic alterations from FFPE slides to a broad range of tumor types, beyond the findings of previous studies which were limited in their selection of genetic alterations (Kather et al., 2020).

MATERIALS AND METHODS

Ethics Statement

All experiments were conducted in accordance with the Declaration of Helsinki and the International Ethical Guidelines for Biomedical Research Involving Human Subjects. Anonymized scanned whole slide images were retrieved from The Cancer Genome Atlas (TCGA) project through the Genomics Data Commons (GDC) Portal (https://portal.gdc.cancer.gov/).

Patient Cohorts

Digitized hematoxylin/Eosin (H and E) stained slides and molecular data from all solid tumor types with more than 100 cases in the GDC database were included in the analysis: bladder urothelial carcinoma [BLCA, *n* = 332 patients, (Robertson et al., 2017)], breast cancer [BRCA, *n* = 977, (Cancer Genome Atlas Network, 2012b)], cervical cancer [CESC, *n* = 253, (Cancer Genome Atlas Research Network et al., 2017a)], colorectal cancer [COAD and READ, merged as CRC, *n* = 499, (Cancer Genome Atlas Network, 2012a)], esophageal cancer [ESCA, *n* = 153, (Cancer Genome Atlas Research Network et al., 2017b)], glioblastoma [GBM, *n* = 200, (Brennan et al., 2013)], head and neck squamous cell carcinoma [HNSC, *n* = 429, (Cancer Genome Atlas Network, 2015a)], clear cell renal cell carcinoma [KIRC, *n* = 376, (Cancer Genome Atlas Research Network, 2013)], papillary renal cell carcinoma [KIRP, *n* = 240, (Cancer Genome Atlas Research Network et al., 2016)], low grade glioma [LGG, *n* = 480, (Cancer Genome Atlas Research Network et al., 2015)], hepatocellular carcinoma [LIHC, *n* = 352, (Cancer Genome Atlas Research Network, 2017b)], lung adenocarcinoma [LUAD, *n* = 457, (Cancer Genome Atlas Research Network, 2014b)], lung squamous cell carcinoma [LUSC, *n* = 410, (Cancer Genome Atlas Research Network, 2012)], ovarian cancer [OV, *n* = 97 after exclusion of non-analyzable samples; nine patients were excluded during analysis, (Cancer Genome Atlas Research Network, 2011)], pancreatic cancer [PAAD, *n* = 166, (Cancer Genome Atlas Research Network, 2017c)], pheochromocytoma and paraganglioma [PCPG, *n* = 169, (Fishbein et al., 2017)], prostate adenocarcinoma [PRAD, *n* = 397, (Cancer Genome Atlas Research Network, 2015)], sarcoma [SARC, *n* = 247, (Cancer Genome Atlas Research Network, 2017a)], melanoma primary tumors [SKCM, *n* = 72, (Cancer Genome Atlas Network, 2015b)] and melanoma metastases [SKCM-M, *n* = 136], gastric cancer [STAD, *n* = 318, (Cancer Genome Atlas Research Network, 2014a)], thymoma [THYM, *n* = 120, (Radovich et al., 2018)], papillary thyroid cancer [THCA, *n* = 479, (Cancer Genome Atlas Research Network, 2014c)], endometrial carcinoma [UCEC, *n* = 470, (Cancer Genome Atlas Research Network et al., 2013)]. In total, 23 solid tumor types with more than 100 patients per tumor type were included. Ten of them were adenocarcinomas (UCEC, CRC, STAD, BRCA, LIHC, THCA, PRAD, LUAD, PAAD, OV), four were mainly squamous cell carcinomas (HNSC, LUSC, CESC, ESCA) and nine were other tumor types, so neither adeno carcinoma or squamous cell carcinoma (LGG, KIRP, GBM, KIRC, BLCA, SARC, PCPG, SKCM, THYM). Although the total patient number was higher than 100, Germ Cell Tumor (TGCT) was not analyzed because this dataset included a wide variety of tumor differentiation patterns, with less than 100 cases per tumor type. Slides from 7,829 patients from the TCGA archive were all from formalin-fixed paraffin-embedded (FFPE) samples.

Image Preprocessing

Regions with tumor were manually annotated with QuPath v0.1.2 (Bankhead et al., 2017) by trained observers in every whole slide image (WSI). The non-pathologist observers were initially trained by experienced histopathologists and consulted the histopathologist to resolve difficult cases. Cases were excluded if the image was of poor quality or did not contain any tumor. Subsequently, the tumor regions within whole slide images were tessellated into tiles of 256 × 256 μm^2 at 0.5 μm per pixel. All data was pre-processed according to the "Aachen Protocol for Deep Learning Histology" (Muti et al., 2020).

Experimental Design and Preprocessing of Mutation Data

Mutation data of all cases was obtained from www.cbioportal.org, accessed on 05/17/19. We included all genes with a mutation prevalence above 5% in cancer populations. In order to select a set of clinically relevant genes, the target genes were selected based on the prevalence of mutations in the MSK-IMPACT Clinical Sequencing Cohort (MSKCC) and OnkoKB (https://www.oncokb.org/, accessed on 06/12/19). In total, 69 genes were analyzed (**Figure 1B**): 18 oncogenes, 44 tumor suppressor genes and seven other genes (**Supplementary Figure S1**). We then ran four different experiments as described as follows. **Experiment #1, "single gene predictability experiment":** For each mutation in each gene, we manually checked whether it is a likely predicted oncogenic mutation based on OnkoKB, Cancer Hotspots, 3D Hotspot or My Cancer Genome (accessed via cbioportal). Based on this, each patient was assigned a status for each gene: mutated (mutated, mutation clinically relevant), wild type (not mutated, mutation not clinically relevant) or inconclusive (no data). For each genetic alteration (mutation of a given gene), we subsequently trained a Deep Learning system to distinguish mutated from wild type cases, counting inconclusive cases as wild type (WT) in order to only include mutated genes in the analysis. **Experiment #2, "pathway predictability experiment":** For the analysis of alterations in pathways, the 69 genes were manually assigned to signaling pathways based on a reference publication (Sanchez-Vega et al., 2018). Genes that were not included in the reference publication were manually assigned to pathways based on an additional review of OncoKB (https://www.oncokb.org/), cbioportal (www.cbioportal.org), Gene cards (https://www.genecards.org/, accessed on 06/12/19) and MyCancerGenome (https://www.mycancergenome.org/, accessed on 06/12/19) databases, literature and expert opinion (**Supplementary Table S1**). In total, 59 genes could be assigned to 12 pathways (**Supplementary Figure S1**): *MAPK, p53, PI3K, Cell cycle, TGFbeta, Hippo, Notch, FOXA1/ESR1, SWI/SNF complex, Jak-STAT, Wnt, Histone Methylation*. The remaining 10 genes could not be assigned to a particular pathway and grouped as "unknown" pathway. Whenever at least one gene assigned to a particular pathway was found to be mutated, the whole pathway was classified as mutated ("pathway-altered") in the tumor; whenever none of the genes were mutated, the pathway was labelled as wild type ("not pathway altered"). For each pathway in each tumor type, we then trained the Deep Learning network to distinguish tumors with altered from those with non-altered pathways (**Figure 1B**). **Experiment #3, "pathway predictability experiment with exclusion of dominant genes":** In addition, we investigated if the predictability of alterations in the pathway was only driven by alterations in a small set of "dominant" genes. To do this, the prediction experiments of pathway-alterations were repeated for three tumor types (UCEC, STAD and CRC) for three pathways (*p53, MAPK, PI3K*), excluding the following genes: *TP53* in *p53* pathway, *BRAF* and *KRAS* in *MAPK* pathway, and *PIK3CA* and *PTEN* for the *PI3K* pathway. The aim of this experiment was to investigate if the predictability of alterations in pathways is driven by alterations in one or two genes or by alterations in a larger set of genes. **Experiment #4, "allele frequency experiment":** Lastly, we performed a correlation analysis between the Deep Learning patient scores and allele frequency for the genes *KRAS* and *TP53* genes across all tumor types.

Deep Learning and Statistics

The general aim of our study was to predict the status of binary targets (single gene mutations or pathway alterations present versus absent) directly from H&E-stained histology image data by Deep Learning. We trained a modified shufflenet for every target as described before (Kather et al., 2020). For each target, the cohort was randomly split into three parts in a stratified way, preserving the proportions of each target level (mutated or wild type). Then, the Deep Learning network was trained in a 3-fold cross-validation approach on the level of patients, ensuring that no image tiles from the same patient were ever part of the training and test set at the same time. Image tiles were only generated from manually annotated tumor regions. Once trained on all tiles in the training set, the network was used to predict the target in each test set tile. Tile-level predictions were subsequently aggregated on the level of patients by simple majority vote and classifier performance was evaluated with a receiver operating curve with 10x bootstrapped 95% pointwise confidence intervals. The primary statistical endpoint was the patient-wise area under the receiver operating curve (AUROC) for each target in each patient cohort. The patient-level prediction scores between patients in the wild type and mutated group for each target were compared by a two-tailed unpaired *t*-test to assess the significance of the separation of groups based on the Deep Learning system. Additionally, for all targets, confusion matrices, F-Score and Matthew correlation coefficient (MCC) with a patient level prediction threshold of 0.5 were calculated and are available in (**Supplementary Figure S2** and **Supplementary Table S2**). Only genes or pathways with at least four patients in each group were analyzed. All source codes are publicly available at https://github.com/jnkather/DeepHistology. A re-implementation of these Matlab codes in Python is available in the histology image analysis package HIA at https://github.com/KatherLab/HIA. All raw histopathology images are available at the TCGA data portal https://portal.gdc.cancer.gov/. All genetic data are available at http://www.cbioportal.org.

RESULTS

Prediction of Clinically Relevant Mutations Directly From Histology

First, we performed a comprehensive screen for the predictability of single gene mutations in the tumor types with more than 100 cases in the GDC database ($n = 23$ tumor types, experiment #1). We systematically tested whether the mutation status of the preselected 69 genes with potential clinical relevance with a mutation prevalence above 5% according to the MSKCC and OnkoKB database is directly predictable from histology slides (a list of all genes and prevalence of their mutations in the analyzed data sets is shown in **Supplementary Figure S3A** and **Supplementary Table S3**). We found that mutations in 44 out of 69 genes were detectable in one or more tumor types. Most consistently, mutations in *TP53* were predictable in 11 out of 23 cohorts (**Figure 2A**) with an average AUROC of 0.6812, ranging from 0.597 in hepatocellular carcinoma (LIHC) (0.5320130.677, $p = 0.035$) to 0.787 (0.758–0.823, $p < 0.001$) in low grade glioma (LGG). In addition, in four of the 23 tumor types, alterations in *PTEN, SETD2* and *KRAS* were identified. *PTEN* prediction reached AUROCs of up to 0.773 (0.73–0.799, $p < 0.001$) in UCEC and 0.773 (0.684–0.826, $p = 0.008$) in BLCA. *SETD2* prediction yielded AUROCs of 0.895 (0.827–0.951, $p = 0.035$) in *PRAD*. *KRAS* mutations were predictable with an AUROC of

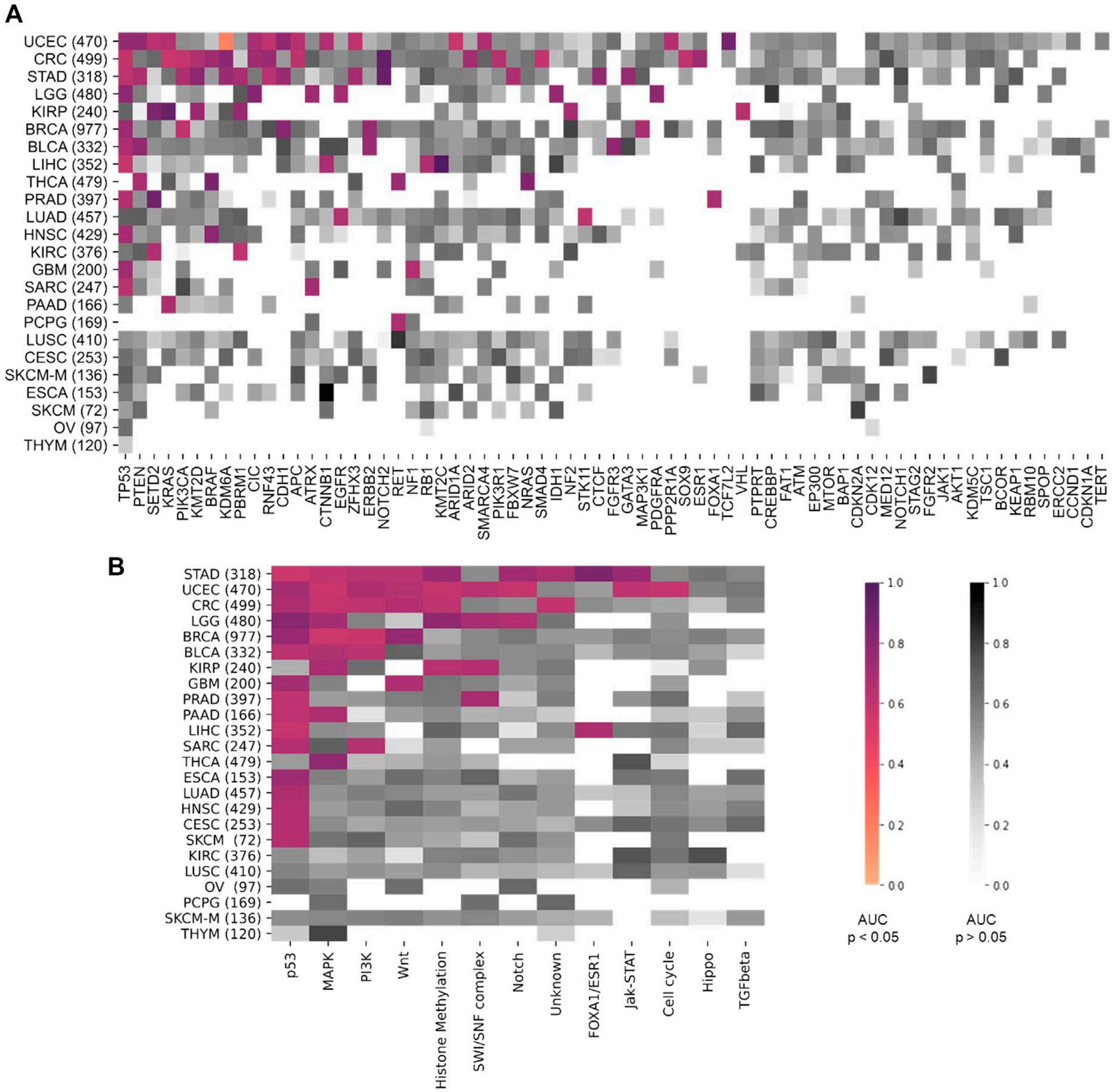


FIGURE 2 | Heatmap comparing the area under the receiver operating curve (AUROC) between the different tumor types. On the y-axis all tumor types are listed and sorted by tumor with most significant results from top to bottom. Number of patients indicated in brackets behind. Pathways are ordered on the x-axis from most (left) to least (right) significant results. AUROC values for **(A)** the twelve pathway analysis and **(B)** for the 69 gene analysis. Coloured values stand for significant detected (*p* > 0.05) pathways and grey for not significantly (*p* > 0.05). TCGA tumor type abbreviations are used (https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations).

0.918 (0.844–0.979, $p < 0.001$) in KIRP. The tumor type with consistently highest AUROCs was UCEC, in which AUROCs of 0.764 (0.694–0.8, $p < 0.001$), 0.773 (0.73–0.799, $p < 0.001$), 0.626 (0.527–0.75, $p = 0.017$) and 0.653 (0.595–0.721, $p < 0.001$) were reached for *TP53, PTEN, SETD2* and *KRAS*, respectively. The neural network predicted alterations of twenty genes very well with AUROCs higher than 0.75. Exemplarily seven of these were selected based on expert opinion by a molecular geneticist (NOB), because they were either most clinically relevant, or associated with morphological patterns or prognosis (**Table 1**). Clinically relevant mutated genes that were chosen were as follows: *FGFR3* in BLCA with an AUROC of 0.78 (0.72–0.822, $p < 0.001$), *IDH1* in LGG with 0.764 (0.735–0.805, $p < 0.001$) and *BRAF* in HNSC with 0.79 (0.739–0.977, $p = 0.001$). Gene mutations associated with morphological patterns were: *BRAF* in THCA 0.86 (0.816–0.886, $p < 0.001$) and *E-Cadherin (CDH1)* in BRCA with 0.81 (0.758–0.849, $p < 0.001$). Prognostically significant mutated genes were: *SETD2* in PRAD with an AUROC of 0.895 (0.827–0.951, $p = 0.005$), *PBRM1* in KIRP with the lowest AUROC 0.752 (0.571–0.939, $p = 0.006$) of these all and lastly *NOTCH2* with the highest AUROC's in CRC 0.934 (0.893–0.978, $p < 0.001$) and STAD 0.919 (0.846–0.982, $p <$

TABLE 1 | Top genes result overview. Single gene analysis results with area under the receiver operating curve (AUROC), confidence interval and *p*-Value. Selected genes were very well predicted by the neural network with AUROCs at least above 0.75. **(1–3)** *FGFR3, PBRM1, IDH1* are clinically relevant, **(4–6)** *CDH1, BRAF* is associated with different morphological features and **(7–9)** *SETD2, NOTCH2* have prognostic value.

ID	Tumor type	Gene	AUROC	*p*-Value
1	BLCA	*FGFR3*	0.78 [0.72–0.822]	<0.001
2	LGG	*IDH1*	0.764 [0.735–0.805]	<0.001
3	HNSC	*BRAF*	0.79 [0.739–0.977]	=0.001
4	THCA	*BRAF*	0.86 [0.816–0.886]	<0.001
5	BRCA	*CDH1*	0.81 [0.758–0.849]	<0.001
6	PRAD	*SETD2*	0.895 [0.827–0.951]	=0.005
7	KIRP	*PBRM1*	0.752 [0.571–0.939]	=0.006
8	CRC	*NOTCH2*	0.934 [0.893–0.978]	<0.001
9	STAD	*NOTCH2*	0.919 [0.846–0.982]	<0.001

0.001). All in all, mutations in 25 genes could not be predicted by the neural network in the gene analysis (**Figure 2A**). Since AUROC is susceptible to different group sizes, we also analyzed the F-Score and MCC (**Supplementary Table S2** and **Supplementary Figure S2**), which showed consistent findings with AUROCs. Among the ten highest F-Scores, *TP53* was found four times. F-scores ranged from 0.846 for *IDH1* in LGG to 0.875 for *TP53* in ESCA and highest MCC correlation of 0.612 for *BRAF* in THCA. In a further analysis, we examined how the F-score, MCC and AUROC changed with different numbers of patients (half $n = 235$, third $n = 157$ and quarter $n = 117$) for the genes *KRAS, PTEN* and *TP53* in the tumor type UCEC ($n = 470$) (**Supplementary Figures S4A–C**). For all values, a decreasing trend was seen with a decreasing number of patients. This effect was strongest for *TP53*.

Prediction of Pathway Alterations Directly From Histology

Next, we tested whether Deep Learning can predict alterations at the level of the selected twelve signaling pathways more easily than the level of individual genes (experiment #2). In this

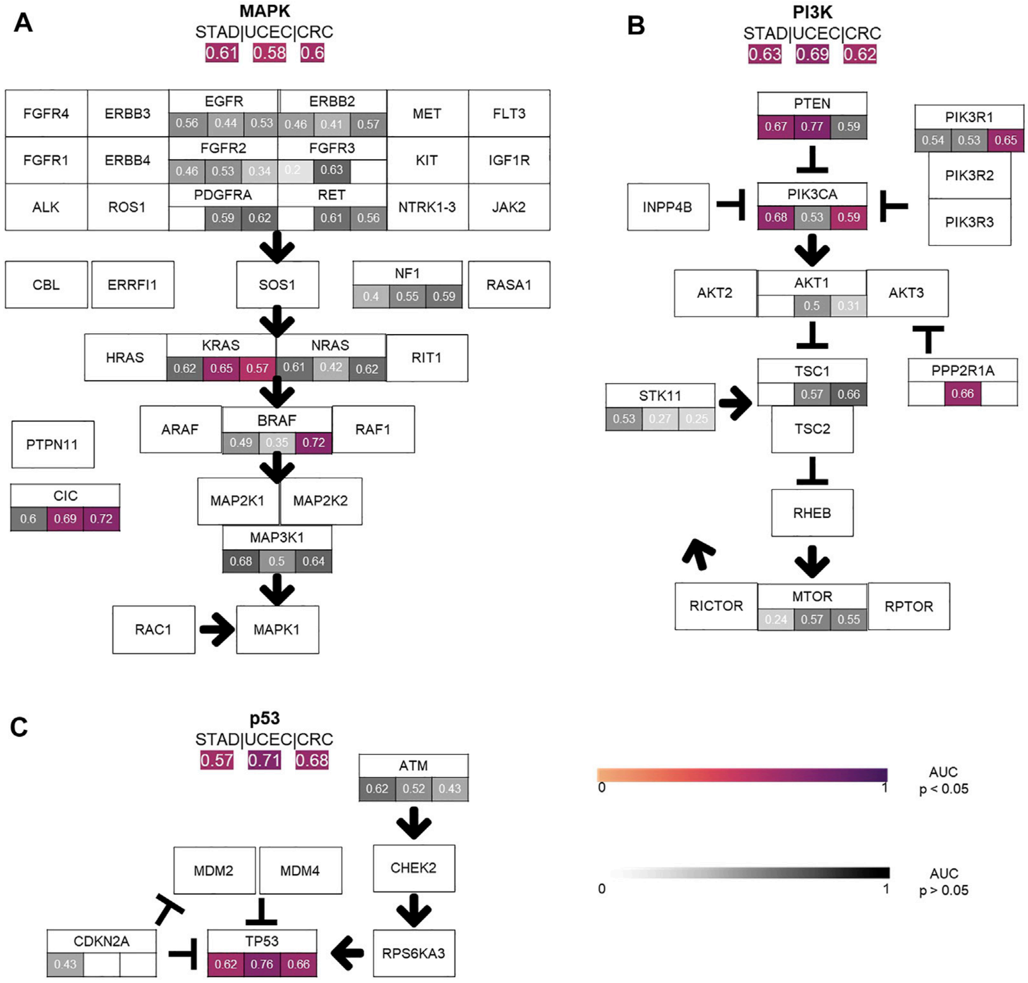


FIGURE 3 | Comparison of the performance of the single gene area under the receiver operating curve (AUROC) vs. pathway AUROC. The top three pathways **(A)** MAPK, **(B)** PI3K, **(C)** TP53 AUROC results for the three top tumor cohorts (STAD, CRC, UCEC) are illustrated. AUROC values are compared between single gene vs. whole pathway. Coloured values stand for significantly detected pathways and grey for not significantly detected ($p > 0.05$).

experiment, a pathway in a given tumor was defined to be altered if at least one of the genes in this pathway were mutated (**Supplementary Figure S3B**). We found that alterations in the pathways *p53, MAPK, PI3K* and *Wnt* were mostly identified by the neural network. The highest number of altered pathways were predictable in gastric cancer (STAD, $n = 318$), endometrial cancer (UCEC, $n = 470$) and colorectal cancer (CRC, $n = 499$) (**Figure 2B**). In many cases, the detection AUROC values for altered genes were similar to those for altered pathways, e.g., for *TP53* detection 0.66 (0.627–0.718, $p < 0.001$) and the for *p53* pathway detection 0.682 (0.668–0.698, $p < 0.001$) in CRC or for *TP53* detection 0.764 (0.694–0.8, $p < 0.001$) and for *p53* pathway detection 0.71 (0.677–0.738, $p < 0.001$) in UCEC. In summary, the AUROCs for altered pathways were in general lower than for individual altered genes (**Figures 2**, **3**) and training on pathway alterations instead of single gene alterations did not consistently yield a higher performance in the 23 cohorts that we analyzed. Based on these data we hypothesized that predictability of alterations in pathways could be primarily driven by the presence of mutations in one or two genes. To address this, we repeated the analysis for predictability of pathway alterations, but excluded the best predictive genes (experiment #3). In this analysis, alterations in the three pathways *p53, MAPK* and *PI3K* could not be significantly predicted, except in the *MAPK* pathway in the STAD cohort with an AUROC 0.633 (0.609–0.699, $p = 0.013$). The predictability of single gene or pathway alterations showed a positive correlation with the absolute number of mutated cases in a given cohort. In BLCA, the status of the genes *PTEN* (MUT = 12, WT = 398), *ERBB2* (MUT = 39, WT = 371) and *FGFR3* (MUT = 64, WT = 346) were all detected with high AUROCs of 0.773 (0.684–0.826, $p = 0.008$), 0.747 (0.64–0.837, $p < 0.001$) and 0.78 (0.72–0.822, $p < 0.001$), respectively. In the BRCA cohort, the status of the genes *CDH1* (MUT = 106, WT = 871), *TP53* (MUT = 311, WT = 666), *MAP3K1* (MUT = 66, WT = 911), *ERBB2* (MUT = 17, WT = 960) and *PIK3CA* (MUT = 312, WT = 661) were significantly ($p < 0.05$) detected with AUROCs above 0.611. The following tumor types had the highest number of predictable genes, and also highest patient numbers: 470 (UCEC), 499 (CRC) and 318 (STAD), while the lowest predictability was seen in tumor types with 97 (OV), 233 (SKCM primary and metastasis) and 120 (THYM) patients in this cohort. However, this relationship was not absolute as for example in KIRP (240 patients), more single gene alterations and pathway alterations were predictable than in LUSC (410 patients). Therefore, we conclude that patient number in a given cohort does not explain the predictability of mutations alone. While alterations of almost all pathways were detectable in one of the tested tumor types, alterations in the pathways *TGF beta* and *Hippo* were not significantly predictable from histology in any tumor type. However, alterations in the gene *SMAD4* could be predicted with an AUROC of 0.601 (0.524–0.669, $p = 0.045$) and likewise mutations of the *NF2* gene reached an AUROC of 0.701 (0.522–0.834, $p = 0.029$). Furthermore, we hypothesized that the predictability of the histological phenotype of a given alteration would correlate with the allele frequency of mutated genes. To test this, we assessed the correlation between patient-level Deep Learning scores and the allele frequency for the genes *TP53* and *KRAS* in all cohorts. However, this analysis failed to demonstrate a significant correlation (**Supplementary Table S4, experiment #4**).

Predictability of Alterations in Different Tumor Types

Having trained Deep Learning systems to detect single gene and pathway alterations in solid tumors, we investigated how tumor types differ in terms of predictability of these alterations. Out of all 23 different tumor types, only in six tumors ($n = 1$ LUSC, $n = 1$ CESC, $n = 1$ SKCM, $n = 1$ ESCA, $n = 1$ OV and $n = 1$ THYM) no mutations were detected. Most altered genes were detected in the cohorts UCEC (15), CRC (15) and STAD (13), all adenocarcinomas. In general, alteration of genes and pathways were identified in nine out of ten adenocarcinoma cohorts, (90%). Three out of the four cohorts of squamous cell carcinomas did not show any significant results. Similar results were seen for the pathway analysis: Most pathway alterations were identified in STAD (9), UCEC (9) and CRC (6). All results are available in **Supplementary Tables S5, 6**.

DISCUSSION

For more than a century, histopathological tissue slides stained with H&E have been the gold standard to diagnose solid tumors. In 2018, a seminal study showed that these images are not only a valuable resource for tumor diagnosis, but that genetic alterations in clinically relevant driver genes an be detected by Deep Learning in lung cancer (Coudray et al., 2018). In 2018 to 2021, a number of studies extended these findings to other tumor types and a wide range of genetic alterations (Couture et al., 2018; Sha et al., 2019; Sun et al., 2019; Zhang et al., 2019; Echle et al., 2020a). In particular in 2020, multiple studies have applied supervised Deep Learning for pan-cancer detection of genetic alterations from snap-frozen samples (Fu et al., 2020; Kather et al., 2020; Schmauch et al., 2020) of the TCGA database. While in this previous study, only a subset of all available tumor types was analyzed, we have now extended the assessment of Deep Learning-based detection of pan-cancer genetic alterations to a wider range of tumor types (from 14 to 23) and observed high detection rates for some clinically interesting genes. Additionally, we have evaluated our Deep Learning approach on pathway level in comparison to focussing on single gene alterations, which has not been tested in previous studies to our knowledge.

We found that alterations in single genes were often better predictable from histology than pathway alterations, suggesting that the phenotypic footprint of a pathway is mostly driven by one or two of the genes and that it might be the gene alteration that creates a recognizable pattern, not the pathway alterations itself (**Figures 3A–C**). The *MAPK* pathway, for example, consists of twelve genes, of whom only three were significantly identified in two cohorts (**Figure 3A**). This can also be seen in the *PI3K* pathway, where mutations in only two out of eight altered genes were significantly detected in gastric, colorectal and endometrial

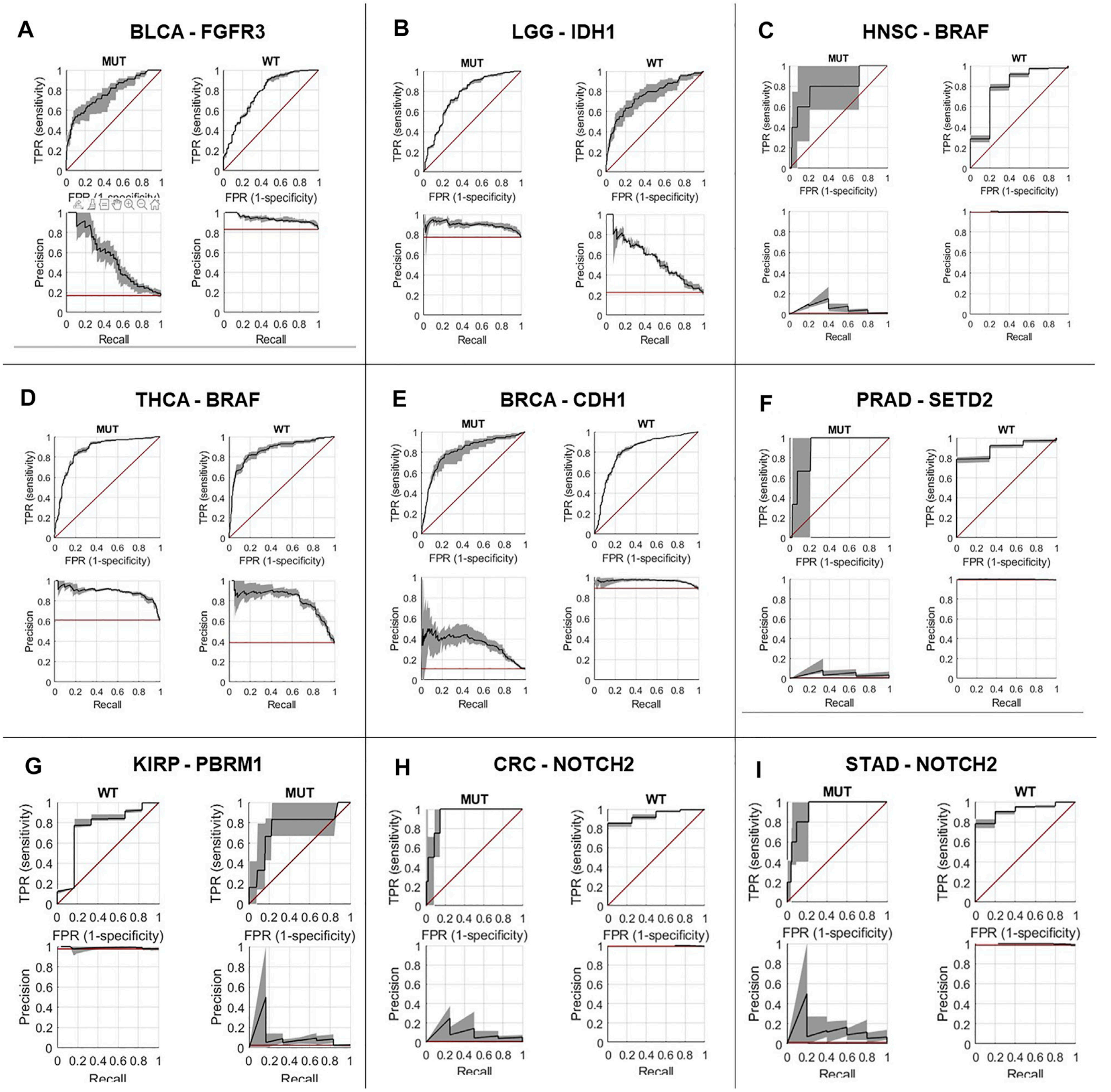


FIGURE 4 | Prediction performance for single gene alterations, representative genes in nine tumor types. Receiver operating curve for: **(A)** *FGFR3* alterations in bladder cancer (BLCA), **(B)** *IDH1* alterations in low grade glioma (LGG), **(C)** *BRAF* alterations in head and neck squamous cell carcinoma (HNSC), **(D)** *BRAF* alterations in thyroid carcinoma (THCA), **(E)** *CDH1* alterations in invasive breast carcinoma (BRCA), **(F)** *SETD2* alterations in prostate adenocarcinoma (PRAD), **(G)** *PBRM1* alterations in renal cell carcinoma (KIRP), **(H)** *NOTCH2* alterations in colorectal adenocarcinoma (CRC), **(I)** *NOTCH2* alterations in stomach adenocarcinoma (STAD), MUT: mutated, WT: wild type

cancer (**Figure 3B**). Single gene alteration AUROC's were similar to those found for altered pathways, e.g., *p53, MAPK* and *PI3K* (**Figure 3C**). This suggests that a commonly mutated gene might determine the outcome of the pathway analysis in some cases. This hypothesis was verified by our pathway analysis excluding highly predictive genes, as pathway alterations could not be significantly predicted. Interestingly, no gene in the STAD cohort was predicted significantly in the *MAPK* pathway, however the AUROC for the altered pathway as a whole was 0.61 (0.55–0.66, p = 0.006) (**Figure 3A**). Another explanation could also be the higher patient numbers in these tumor cohorts, since this also influences the predictability of alterations in genes and pathways (**Figure 3**). This was confirmed by a further exemplary analysis in which the AUROC, F-score and MCC decreased with a smaller number of patients in the cohort UCEC (**Supplementary Figure S4**). Direct prediction of mutated single

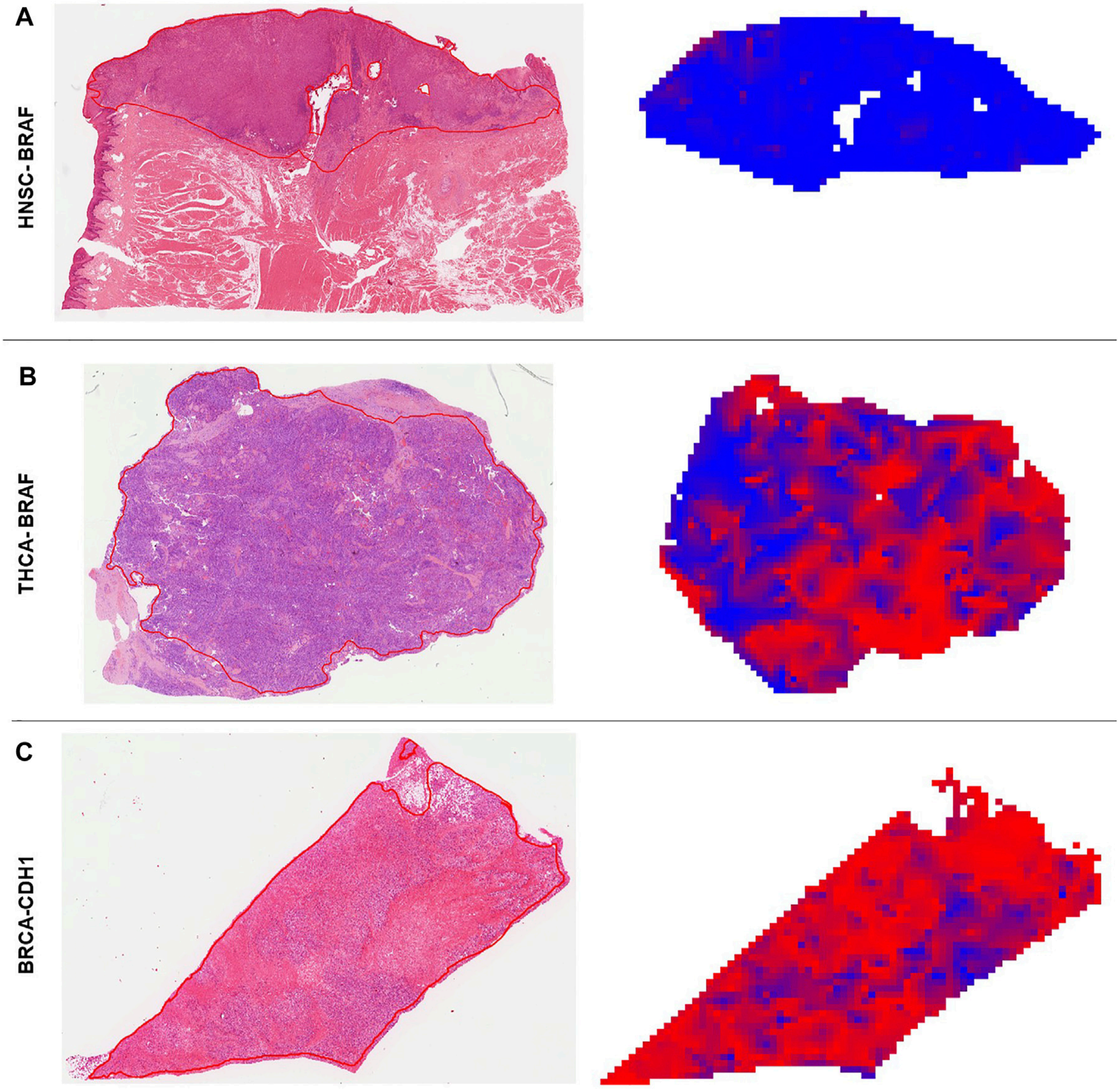


FIGURE 5 | Deep learning predicted heatmaps. Visualization of manually annotated histological slides hematoxylin & eosin (H&E) with corresponding prediction maps for altered genes. Blue areas are wild type (WT) predicted regions and red areas are identified as mutated (MUT) parts by the neural network. **(A)** H&E slide of a *BRAF* WT patient (ID: TCGA-CQ-5333) from the head and neck squamous cell carcinoma (HNSC) cohort. The homogenous blue heatmap is consistent with the wild type status of the patient. **(B)** H&E slide of a *BRAF* mutated patient (ID: TCGA-EL-A3H7) from the thyroid carcinoma (THCA) cohort. The heatmap is more than 50% red, which means the patient was correctly classified as MUT. Intermingled blue areas in tumor regions reflect stroma and artifacts that disturb these areas. **(C)** H&E slide of a *CDH1* mutated patient (ID: TCGA-PE-A5DD) of the breast invasive carcinoma (BRCA) cohort. The prediction heatmap shows that stroma tissue is mostly predicted as WT (blue areas condensed connective tissue) and diffuse invasive-lobular cancer is mostly red.

genes from histology images is potentially useful, especially if the alterations have a clinical implication. In general, the neural network could predict several genes—some of which are clinically relevant, associated with morphological pattern or prognostically relevant—very well with AUROCs higher than 0.75. For example, in our study, *FGFR3* mutations could be predicted with an AUROC of 0.78 in bladder cancer (**Figure 4A**). Since the FDA approved the first targeted-therapy with the *FGFR* inhibitor erdafitinib in advanced muscle invasive bladder cancer (U.S. Food and Drug Administration, 2022), detection of *FGFR3* could identify patients who might benefit from this therapy (Loriot et al., 2019). *IDH1* is an important prognostic marker for brain tumors (Young et al., 2020). In the LGG cohort, 77% (395/512) were *IDH1* mutated, which is associated with a better outcome. *IDH1* could be detected significantly in LGG with an AUROC of 0.764 (**Figure 4B**). However, in GBM, where only 6% of tumors were mutated, *IDH1* was not significantly detectable. Effectiveness of *IDH* specific enzyme inhibitors in brain tumors are currently tested in clinical trials (Karpel-Massler et al., 2019).

The V600E mutation of the *BRAF* kinase gene, which is part of the *MAPK* pathway, plays an important role in tumorigenesis across many types of solid tumors and is in fact a highest level evidence gene of OnkoKB. Although the *MAPK* pathway is altered in many tumor types, *BRAF* mutations are not very common in HNSC (Weber et al., 2003), including the TCGA-HNSC cohort, where only five out of 515 patients showed a *BRAF* alteration. Still, *BRAF* alterations were recognized with a performance of 0.79 in our analysis (**Figure 4C** and **Figure 5A**). This makes Deep Learning based identification of subgroups that might benefit from targeted therapy in HNSC conceivable, as specific *BRAF* and *MEK* inhibitors are already an integral part of guideline-directed therapy in other entities. However, interestingly most of the detected mutations in HNSC were indeed non-V600 class II or class III mutations (Yaeger and Corcoran, 2019). In contrast, more than 40% of the thyroid cancers show a BRAF V600 mutation which is associated with the papillary tumor type and found rarely in follicular thyroid cancer type (Nikiforov, 2011) (**Figure 5B**). In fact, *BRAF* mutational status could be predicted significantly with an AUROC of 0.86 (**Figure 4D**). Another example is *E-cadherin*, a tumor suppressor gene which is mostly involved in cell adhesion, which is associated with the lobular subtype of breast cancer (Cancer Genome Atlas Network, 2012b) (**Figure 5C**). In our analysis it was detectable with an AUROC of 0.8 (**Figure 4E**). Another example is the *PBRM1* gene, which belongs to the *SWI/SNF* chromatin remodelling complex. *PBRM1* is often altered in renal papillary carcinoma, however, recent studies have shown that a *PBRM1* mutation correlates with decreased survival (Ricketts et al., 2018). In our analysis, *PBRM1* was identified with an AUROC of 0.752 (**Figure 4F**). As *PBRM1* alterations are not very common in papillary renal cell carcinoma, its significance in terms of a potential biomarker remains to be elucidated (Ho et al., 2015; Liu et al., 2020). In KIRP, loss of *PBRM1* has been described to be associated with checkpoint inhibitor resistance (Ho et al., 2015; Liu et al., 2020). Highly predictive image tiles of the five genes mentioned above are collected in **Supplementary Figure S5**. Other significantly detected prognostic alterations in genes were found in *SETD2* in PRAD with an AUROC of 0.895 (Yuan et al., 2020) and *NOTCH2* in CRC and STAD (Chu et al., 2011) with an high AUROC of 0.934, 0.919 (**Figures 4G–I**). However, these two genes did not show any relevant pathological features in the top tiles analysis. Analysis of phenotypic footprint of an alteration did not correlate with the allele frequency of mutated genes (**Supplementary Table S4**).

Based on our overall results, genetic alterations in adenocarcinomas were better predictable than alterations in other tumor types, such as squamous cell carcinomas (**Figure 2**). This is consistent with previous studies (Schmauch et al., 2020) and leads us to hypothesize that in tumor types with glandular architecture genetic changes might more frequently result in morphological changes and therefore better detectable. Some tumors, predominantly UCEC, CRC and STAD, had more numerous significant findings than others. Interestingly, these are not the tumors with high mutational burden (Kandoth et al., 2013).

In summary, H&E stained tumor images contain subtle morphogenetic information which is detectable by Deep Learning. Our findings correlate with similar results of other Deep Learning analyses and mutational landscape across cancers (Kandoth et al., 2013; Fu et al., 2020; Kather et al., 2020).

Limitations

Our study has a number of limitations. first and foremost the use of TCGA as our only resource for histopathological whole slide images, which means that validation on additional cohorts is necessary to confirm the results. In addition, a potential confounder in our study is the unequal dataset size for different tumor types in TCGA. It is possible and likely that our study underestimated the number of predictable genes in some tumor types. Especially for small cohorts in this study, future studies should re-analyze the same set of genes in larger cohorts, once such cohorts become available. Finally, while this archive is undoubtedly the most comprehensive multicentric resource available to computational pathology researchers, it has been shown to carry a risk of bias due to the patient selection process in TCGA (Howard et al., 2020). However, a full genetic characterization of thousands of tumor samples like in TCGA is an almost impossible task for academic research groups, which is why TCGA remains very useful and unique to develop and test new computational pathology approaches. Yet, even the genomic characterization in TCGA carries some ambiguity, e.g., due to the presence of non-tumor tissue in sequenced samples as well as different methods for mutation calling. We focussed on single nucleotide variants and small deletions/insertions, and did not take into account fusion genes, copy number changes or expression data. We also relied on a conservative variant classification and therefore might have created a bias regarding the inclusion of "false negative" samples. In future studies, it could be interesting not only to include clinical variant classification data but instead also narrow down the number of included unclassified variants by using prediction algorithms as e.g., BoostDM (https://www.intogen.org/boostdm/search). The most promising candidates for clinical translation should be evaluated in other multicentric image collections obtained via academic consortia. Another limitation of our study is that the tissue slides which we used for our prediction do not necessarily contain the same region that the DNA for genetic characterization has been extracted from. Therefore, it is conceivable that intratumor heterogeneity could dilute our results, potentially leading to a lower performance. Further studies are needed to systematically quantify the impact of intratumor genetic heterogeneity on the inference of genetic alterations from pathology images.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/.

AUTHOR CONTRIBUTIONS

CL, NB, and JK designed the study. CL performed the experiments and wrote the first draft of the manuscript. NL, JK, and MV, contributed to the data analysis. HM, AE, CT, LH, and HG contributed to the interpretation of the results. NG and HG provided histopathology expertise. All authors contributed to the final manuscript and approved the submission.

FUNDING

JK is funded by the Max-Eder-Programme of the German Cancer Aid (Bonn, Germany; grant #70113864) and the German Ministry of Health ("Förderung aufgrund eines Beschlusses des Deutschen Bundestages durch die Bundesregierung"; grant DEEP LIVER, #ZMVI1-2520DAT111). NG is funded by the German Research Foundation (Bonn, Germany; grant GA 1384/3-1). CT is supported by the German Research Foundation (DFG) (SFB CRC1382, SFB-TRR57).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.806386/full#supplementary-material

Supplementary Figure S1 | Overview of the genes and pathways arrangement. Following 59 genes were grouped into twelve different cancer pathways. Legend shows characteristics of the genes (orange, oncogenes; blue, tumor suppressor genes; green, both). Information was taken from OnkoKB (https://oncokb.org/).

Supplementary Figure S2 | Diagram of F-Score and Matthews Correlation Coefficient vs. AUROC. **(A)** Shown are the F-Score values vs. area under the receiver operating curve (AUROC) for each investigated target of all 23 tumours. **(B)** Shown are the Matthews Correlation Coefficient (MCC) values vs. AUROC for each investigated target of all 23 tumours.

Supplementary Figure S3 | Prevalence of mutations in genes and pathways in all tumor types. **(A)** Mutation prevalence in the analyzed data set. **(B)** Prevalence of pathway alterations in the analyzed data set.

Supplementary Figure S4 | Diagram AUROC, F-Score and Matthew Correlation Coefficient (MCC) for different patient numbers. **(A)** Changes in area under the receiver operating curve, F-Score and MCC with half, third and fourth patient numbers in tumor type UCEC for target *KRAS*. **(B)** Changes in area under the receiver operating curve, F-Score and MCC with half, third and fourth patient numbers in tumor type UCEC for target *PTEN*. **(C)** Changes in area under the receiver operating curve, F-Score and MCC with half, third and fourth patient numbers in tumor type UCEC for target *TP53*.

Supplementary Figure S5 | Highly predictive image tiles selected by deep learning. Comparison of the highly ranked mutated (MUT) and wild type (WT) tiles. This visualization helps to identify morphological changes due to alterations of the tumor. **(A)** *FGFR3* in BLCA: MUT top tiles exhibit more papillary structured parts, whereas WT tiles are more diffuse infiltrative tumor parts. **(B)** *PBRM1* in KIRP: MUT tiles are more solid with rosette-like arrangements, whereas WT tiles have a papillary architecture. **(C)** *IDH1* in LGG: MUT tiles show more glial and fibril appearance. In contrast, WT tiles have a higher nuclear density. **(D)** *CDH1* in BRCA: MUT tiles show the diffusely infiltrating, indian file pattern of lobular-invasive breast cancer, while WT tiles have the trabecular, nodular or tubular architecture of no-special type carcinomas. **(E)** *BRAF* in HNSC: MUT tiles are high grade cancers without any squamoid/cornified elements, whereas in the WT tiles squamous cells and cornification can be found. **(F)** *BRAF* in THCA: MUT of *BRAF* is associated with papillary tumor types and WT tiles show a more organ specific follicular/colloidal histology.

Supplementary Table S1 | Raw results (Microsoft Excel file) of all genes and grouping based on different data sources.

Supplementary Table S2 | Confusion matrices, F-Score and Matthews Correlation Coefficient (MCC) for each target, based on a patient-level cutoff of 0.5.

Supplementary Table S3 | Mutation prevalences for pathway and gene analysis.

Supplementary Table S4 | Correlation analysis of allele frequency with mutation score for *TP53* and *KRAS*.

Supplementary Table S5 | Excel File for all AUROC combined Pathway and Genes.

Supplementary Table S6 | Prediction results for all tumor types with all statistics. AUROC_avg, average area under the receiver operating curve (AUROC); AUROC_low, lower bound of AUROC confidence interval (CI); AUROC_high, upper bound of AUROC CI; AUCPR_avg, average area under the precision recall curve (AUCPR); AUCPR_low, lower bound of AUCPR CI; AUCPR_high, upper bound of AUCPR CI.; meanCat, mean prediction values for patients in this category; meanOth, mean prediction values for other patients; pVal, p-value for comparison of prediction scores between patients in target category vs. not in target category (according to the ground truth). nPatsTotal, total number of patients (sum of patients in category and not in category).

REFERENCES

Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., et al. (2017). QuPath: Open Source Software for Digital Pathology Image Analysis. *Sci. Rep.* 7, 16878. doi:10.1038/s41598-017-17204-5

Ben-Hamo, R., Jacob Berger, A., Gavert, N., Miller, M., Pines, G., Oren, R., et al. (2020). Predicting and Affecting Response to Cancer Therapy Based on Pathway-Level Biomarkers. *Nat. Commun.* 11, 3296. doi:10.1038/s41467-020-17090-y

Cancer Genome Atlas Research NetworkBrat, D. J., Brat, D. J., Verhaak, R. G., Aldape, K. D., Yung, W. K., Salama, S. R., et al. (2015). Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N. Engl. J. Med.* 372, 2481–2498. doi:10.1056/NEJMoa1402121

Brennan, C. W., Verhaak, R. G., McKenna, A., Campos, B., Noushmehr, H., Salama, S. R., et al. (2013). The Somatic Genomic Landscape of Glioblastoma. *Cell* 155, 462–477. doi:10.1016/j.cell.2013.09.034

Cancer Genome Atlas Network (2015a). Comprehensive Genomic Characterization of Head and Neck Squamous Cell Carcinomas. *Nature* 517, 576–582. doi:10.1038/nature14129

Cancer Genome Atlas Network (2012a). Comprehensive Molecular Characterization of Human colon and Rectal Cancer. *Nature* 487, 330–337. doi:10.1038/nature11252

Cancer Genome Atlas Network (2012b). Comprehensive Molecular Portraits of Human Breast Tumours. *Nature* 490, 61–70. doi:10.1038/nature11412

Cancer Genome Atlas Network (2015b). Genomic Classification of Cutaneous Melanoma. *Cell* 161, 1681–1696. doi:10.1016/j.cell.2015.05.044

Cancer Genome Atlas Research Network (2017a). Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell* 171, 950–e28. doi:10.1016/j.cell.2017.10.014

Cancer Genome Atlas Research Network (2017b). Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* 169, 1327–e23. doi:10.1016/j.cell.2017.05.046

Cancer Genome Atlas Research Network (2012). Comprehensive Genomic Characterization of Squamous Cell Lung Cancers. *Nature* 489, 519–525. doi:10.1038/nature11404

Cancer Genome Atlas Research Network (2013). Comprehensive Molecular Characterization of clear Cell Renal Cell Carcinoma. *Nature* 499, 43–49. doi:10.1038/nature12222

Cancer Genome Atlas Research Network (2014a). Comprehensive Molecular Characterization of Gastric Adenocarcinoma. *Nature* 513, 202–209. doi:10.1038/nature13480

Cancer Genome Atlas Research Network (2014b). Comprehensive Molecular Profiling of Lung Adenocarcinoma. *Nature* 511, 543–550. doi:10.1038/nature13385

Cancer Genome Atlas Research Network (2011). Integrated Genomic Analyses of Ovarian Carcinoma. *Nature* 474, 609–615. doi:10.1038/nature10166

Cancer Genome Atlas Research Network (2017c). Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* 32, 185–e13. doi:10.1016/j.ccell.2017.07.007

Cancer Genome Atlas Research Network (2014c). Integrated Genomic Characterization of Papillary Thyroid Carcinoma. *Cell* 159, 676–690. doi:10.1016/j.cell.2014.09.050

Cancer Genome Atlas Research Network (2015). The Molecular Taxonomy of Primary Prostate Cancer. *Cell* 163, 1011–1025. doi:10.1016/j.cell.2015.10.025

Chu, D., Zhang, Z., Zhou, Y., Wang, W., Li, Y., Zhang, H., et al. (2011). Notch1 and Notch2 Have Opposite Prognostic Effects on Patients with Colorectal Cancer. *Ann. Oncol.* 22, 2440–2447. doi:10.1093/annonc/mdq776

Coudray, N., Ocampo, P. S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., et al. (2018). Classification and Mutation Prediction from Non-small Cell Lung Cancer Histopathology Images Using Deep Learning. *Nat. Med.* 24, 1559–1567. doi:10.1038/s41591-018-0177-5

Couture, H. D., Williams, L. A., Geradts, J., Nyante, S. J., Butler, E. N., Marron, J. S., et al. (2018). Image Analysis with Deep Learning to Predict Breast Cancer Grade, ER Status, Histologic Subtype, and Intrinsic Subtype. *npj Breast Cancer* 4, 30. doi:10.1038/s41523-018-0079-1

Dhillon, A. S., Hagan, S., Rath, O., and Kolch, W. (2007). MAP Kinase Signalling Pathways in Cancer. *Oncogene* 26, 3279–3290. doi:10.1038/sj.onc.1210421

Echle, A., Grabsch, H. I., Quirke, P., van den Brandt, P. A., West, N. P., Hutchins, G. G. A., et al. (2020a). Clinical-Grade Detection of Microsatellite Instability in Colorectal Tumors by Deep Learning. *Gastroenterology* 159, 1406–1416. doi:10.1053/j.gastro.2020.06.021

Echle, A., Rindtorff, N. T., Brinker, T. J., Luedde, T., Pearson, A. T., and Kather, J. N. (2020b). Deep Learning in Cancer Pathology: a New Generation of Clinical Biomarkers. *Br. J. Cancer* 124, 686–696. doi:10.1038/s41416-020-01122-x

Cancer Genome Atlas Research Network; Albert Einstein College of Medicine; Analytical Biological Services; Barretos Cancer Hospital; Baylor College of Medicine; Beckman Research Institute of City of Hope, et al. (2017a). Integrated Genomic and Molecular Characterization of Cervical Cancer. *Nature* 543, 378–384. doi:10.1038/nature21386

Cancer Genome Atlas Research Network; Analysis Working Group: Asan University; BC Cancer Agency; Brigham and Women's Hospital; Broad Institute; Brown University, et al. (2017b). Integrated Genomic Characterization of Oesophageal Carcinoma. *Nature* 541, 169–175. doi:10.1038/nature20805

Fishbein, L., Leshchiner, I., Walter, V., Danilova, L., Robertson, A. G., Johnson, A. R., et al. (2017). Comprehensive Molecular Characterization of Pheochromocytoma and Paraganglioma. *Cancer Cell* 31, 181–193. doi:10.1016/j.ccell.2017.01.001

Fu, Y., Jung, A. W., Torne, R. V., Gonzalez, S., Vöhringer, H., Shmatko, A., et al. (2020). Pan-cancer Computational Histopathology Reveals Mutations, Tumor Composition and Prognosis. *Nat. Cancer* 1, 800–810. doi:10.1038/s43018-020-0085-8

Ho, T. H., Kapur, P., Joseph, R. W., Serie, D. J., Eckel-Passow, J. E., Parasramka, M., et al. (2015). Loss of PBRM1 and BAP1 Expression Is Less Common in Non-clear Cell Renal Cell Carcinoma Than in clear Cell Renal Cell Carcinoma. *Urol. Oncol.* 33, 23–e14. doi:10.1016/j.urolonc.2014.10.014

Howard, F. M., Dolezal, J., Kochanny, S., Schulte, J., Chen, H., Heij, L., et al. (2020). The Impact of Digital Histopathology Batch Effect on Deep Learning Model Accuracy and Bias. *Nat Commun 12* 4423. doi:10.1038/s41467-021-24698-1

Jiang, N., Dai, Q., Su, X., Fu, J., Feng, X., and Peng, J. (2020). Role of PI3K/AKT Pathway in Cancer: the Framework of Malignant Behavior. *Mol. Biol. Rep.* 47, 4587–4629. doi:10.1007/s11033-020-05435-1

Cancer Genome Atlas Research NetworkKandoth, C., Kandoth, C., Schultz, N., Cherniack, A. D., Akbani, R., Liu, Y., et al. (2013). Integrated Genomic Characterization of Endometrial Carcinoma. *Nature* 497, 67–73. doi:10.1038/nature12113

Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., et al. (2013). Mutational Landscape and Significance across 12 Major Cancer Types. *Nature* 502, 333–339. doi:10.1038/nature12634

Karpel-Massler, G., Nguyen, T. T. T., Shang, E., and Siegelin, M. D. (2019). Novel IDH1-Targeted Glioma Therapies. *CNS Drugs* 33, 1155–1166. doi:10.1007/s40263-019-00684-6

Kather, J. N., and Calderaro, J. (2020). Development of AI-Based Pathology Biomarkers in Gastrointestinal and Liver Cancer. *Nat. Rev. Gastroenterol. Hepatol.* 17, 591–592. doi:10.1038/s41575-020-0343-3

Kather, J. N., Heij, L. R., Grabsch, H. I., Loeffler, C., Echle, A., Muti, H. S., et al. (2020). Pan-cancer Image-Based Detection of Clinically Actionable Genetic Alterations. *Nat. Cancer* 1, 789–799. doi:10.1038/s43018-020-0087-6

Kather, J. N., Pearson, A. T., Halama, N., Jäger, D., Krause, J., Loosen, S. H., et al. (2019). Deep Learning Can Predict Microsatellite Instability Directly from Histology in Gastrointestinal Cancer. *Nat. Med.* 25, 1054–1056. doi:10.1038/s41591-019-0462-y

Cancer Genome Atlas Research NetworkLinehan, W. M., Linehan, W. M., Spellman, P. T., Ricketts, C. J., Creighton, C. J., Fei, S. S., et al. (2016). Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N. Engl. J. Med.* 374, 135–145. doi:10.1056/NEJMoa1505917

Liu, X.-D., Kong, W., Peterson, C. B., McGrail, D. J., Hoang, A., Zhang, X., et al. (2020). PBRM1 Loss Defines a Nonimmunogenic Tumor Phenotype Associated with Checkpoint Inhibitor Resistance in Renal Carcinoma. *Nat. Commun.* 11, 2135. doi:10.1038/s41467-020-15959-6

Loeffler, C. M. L., Ortiz Bruechle, N., Jung, M., Seillier, L., Rose, M., Laleh, N. G., et al. (2021). Artificial Intelligence-Based Detection of FGFR3 Mutational Status Directly from Routine Histology in Bladder Cancer: A Possible Preselection for Molecular Testing? *Eur. Urol. Focus* S2405-4569, 00113. doi:10.1016/j.euf.2021.04.007

Loriot, Y., Necchi, A., Park, S. H., Garcia-Donas, J., Huddart, R., Burgess, E., et al. (2019). Erdafitinib in Locally Advanced or Metastatic Urothelial Carcinoma. *N. Engl. J. Med.* 381, 338–348. doi:10.1056/nejmoa1817323

Muti, H. S., Heij, L. R., Keller, G., Kohlruss, M., Langer, R., Dislich, B., et al. (2021). Development and Validation of Deep Learning Classifiers to Detect Epstein-Barr Virus and Microsatellite Instability Status in Gastric Cancer: a Retrospective Multicentre Cohort Study. *The Lancet Digital Health* 3, e654. doi:10.1016/s2589-7500(21)00133-3

Muti, H. S., Loeffler, C., Echle, A., Heij, L. R., Buelow, R. D., Krause, J., et al. (2020). *The Aachen Protocol for Deep Learning Histopathology: A Hands-On Guide for Data Preprocessing*. Zenodo: Aachen. doi:10.5281/zenodo.3694994

Naik, N., Madani, A., Esteva, A., Keskar, N. S., Press, M. F., Ruderman, D., et al. (2020). Deep Learning-Enabled Breast Cancer Hormonal Receptor Status Determination from Base-Level H&E Stains. *Nat. Commun.* 11, 5727. doi:10.1038/s41467-020-19334-3

Nikiforov, Y. E. (2011). Molecular Analysis of Thyroid Tumors. *Mod. Pathol.* 24 (Suppl. 2), S34–S43. doi:10.1038/modpathol.2010.167

Radovich, M., Pickering, C. R., Felau, I., Ha, G., Zhang, H., Jo, H., et al. (2018). The Integrated Genomic Landscape of Thymic Epithelial Tumors. *Cancer Cell* 33, 244–e10. doi:10.1016/j.ccell.2018.01.003

Ricketts, C. J., De Cubas, A. A., Fan, H., Smith, C. C., Lang, M., Reznik, E., et al. (2018). The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep* 23, 3698–4326. doi:10.1016/j.celrep.2018.06.032

Robertson, A. G., Kim, J., Al-Ahmadie, H., Bellmunt, J., Guo, G., Cherniack, A. D., et al. (2017). Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell* 171, 540–e25. doi:10.1016/j.cell.2017.09.007

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W. K., Luna, A., La, K. C., et al. (2018). Oncogenic Signaling Pathways in the Cancer Genome Atlas. *Cell* 173, 321–e10. doi:10.1016/j.cell.2018.03.035

Schmauch, B., Romagnoni, A., Pronier, E., Saillard, C., Maillé, P., Calderaro, J., et al. (2020). A Deep Learning Model to Predict RNA-Seq Expression of Tumours from Whole Slide Images. *Nat. Commun.* 11, 3877. doi:10.1038/s41467-020-17678-4

Schumacher, D., Andrieux, G., Boehnke, K., Keil, M., Silvestri, A., Silvestrov, M., et al. (2019). Heterogeneous Pathway Activation and Drug Response Modelled in Colorectal-Tumor-Derived 3D Cultures. *Plos Genet.* 15, e1008076. doi:10.1371/journal.pgen.1008076

Sha, L., Osinski, B. L., Ho, I. Y., Tan, T. L., Willis, C., Weiss, H., et al. (2019). Multi-Field-of-View Deep Learning Model Predicts Nonsmall Cell Lung Cancer Programmed Death-Ligand 1 Status from Whole-Slide Hematoxylin and Eosin Images. *J. Pathol. Inform.* 10, 24. doi:10.4103/jpi.jpi_24_19

Sun, M., Zhou, W., Qi, X., Zhang, G., Girnita, L., Seregard, S., et al. (2019). Prediction of BAP1 Expression in Uveal Melanoma Using Densely-Connected Deep Classification Networks. *Cancers* 11, 1579. doi:10.3390/cancers11101579

U.S. Food and Drug Administration (2022). FDA-approved Drugs. Drugs@FDA: FDA-Approved Drugs. Available at: : https://www.accessdata.fda.gov/scripts/cder/daf/(Accessed August 31, 2021).

Weber, A., Langhanki, L., Sommerer, F., Markwarth, A., Wittekind, C., and Tannapfel, A. (2003). Mutations of the BRAF Gene in Squamous Cell Carcinoma of the Head and Neck. *Oncogene* 22, 4757–4759. doi:10.1038/sj.onc.1206705

Yaeger, R., and Corcoran, R. B. (2019). Targeting Alterations in the RAF-MEK Pathway. *Cancer Discov.* 9, 329–341. doi:10.1158/2159-8290.cd-18-1321

Young, J. S., Gogos, A. J., Morshed, R. A., Hervey-Jumper, S. L., and Berger, M. S. (2020). Molecular Characteristics of Diffuse Lower Grade Gliomas: what Neurosurgeons Need to Know. *Acta Neurochir* 162, 1929–1939. doi:10.1007/s00701-020-04426-2

Yuan, H., Han, Y., Wang, X., Li, N., Liu, Q., Yin, Y., et al. (2020). SETD2 Restricts Prostate Cancer Metastasis by Integrating EZH2 and AMPK Signaling Pathways. *Cancer Cell* 38, 350–365. doi:10.1016/j.ccell.2020.05.022

Zhang, H., Zhang, F., Ren, F., Wang, Z., Rao, X., Li, L., et al. (2019). "Predicting Tumor Mutational Burden from Liver Cancer Pathological Images Using Convolutional Neural Network," in 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), San Diego, CA, USA, 18-21 Nov. 2019. doi:10.1109/bibm47256.2019.8983139

Conflict of Interest: JK declares consulting services for Owkin, France and Panakeia, UK.

The remaining 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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 24 February 2022
doi: 10.3389/fgene.2022.801484

Exploring Immune-Related Prognostic Signatures in the Tumor Microenvironment of Colon Cancer

Lichao Cao[1,2], *Tong Li*[1,2], *Ying Ba*[3], *Erfei Chen*[1,2], *Jin Yang*[1,2]* *and Hezi Zhang*[3]*

[1]Provincial Key Laboratory of Biotechnology of Shaanxi Province, Northwest University, Xi'an, China, [2]Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, School of Life Sciences, Northwest University, Xi'an, China, [3]Shenzhen Nucleus Gene Technology Co., Ltd., Shenzhen, China

Background: Colon cancer is a common malignant tumor with poor prognosis. The aim of this study is to explore the immune-related prognostic signatures and the tumor immune microenvironment of colon cancer.

Methods: The mRNA expression data of TCGA-COAD from the UCSC Xena platform and the list of immune-related genes (IRGs) from the ImmPort database were used to identify immune-related differentially expressed genes (DEGs). Then, we constructed an immune-related risk score prognostic model and validated its predictive performance in the test dataset, the whole dataset, and two independent GEO datasets. In addition, we explored the differences in tumor-infiltrating immune cell types, tumor mutation burden (TMB), microsatellite status, and expression levels of immune checkpoints and their ligands between the high-risk and low-risk score groups. Moreover, the potential value of the identified immune-related signature with respect to immunotherapy was investigated based on an immunotherapeutic cohort (Imvigor210) treated with an anti-PD-L1 agent.

Results: Seven immune-related DEGs were identified as prognostic signatures. The areas under the curves (AUCs) of the constructed risk score model for overall survival (OS) were calculated (training dataset: 0.780 at 3 years, 0.801 at 4 years, and 0.766 at 5 years; test dataset: 0.642 at 3 years, 0.647 at 4 years, and 0.629 at 5 years; and the whole dataset: 0.642 at 3 years, 0.647 at 4 years, and 0.629 at 5 years). In the high-risk score group of the whole dataset, patients had worse OS, higher TMN stages, advanced pathological stages, and a higher *TP53* mutation rate ($p < 0.05$). In addition, a high level of resting NK cells or M0 macrophages, and high TMB were significantly related to poor OS ($p < 0.05$). Also, we observed that high-risk score patients had a high expression level of *PD-L1*, *PD-1*, and *CTLA-4* ($p < 0.05$). The patients with high-risk scores demonstrated worse prognosis than those with low-risk scores in multiple datasets (GSE39582: $p = 0.0023$; GSE17536: $p = 0.0008$; immunotherapeutic cohort without platinum treatment: $p = 0.0014$; immunotherapeutic cohort with platinum treatment: $p = 0.0027$).

Conclusion: We developed a robust immune-related prognostic signature that performed great in multiple cohorts and explored the characteristics of the tumor immune microenvironment of colon cancer patients, which may give suggestions for the prognosis and immunotherapy in the future.

Keywords: colon cancer, tumor immune microenvironment, prognostic model, immunotherapy, TCGA-COAD, GEO

Edited by:
Luis Zapata,
Institute of Cancer Research (ICR),
United Kingdom

Reviewed by:
Matjaz Rokavec,
Ludwig Maximilian University of Munich, Germany
Anil K. Giri,
University of Helsinki, Finland

***Correspondence:**
Jin Yang
yangjin@nwu.edu.cn
Hezi Zhang
zhz358@126.com

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *25 October 2021*
Accepted: *26 January 2022*
Published: *24 February 2022*

Citation:
Cao L, Li T, Ba Y, Chen E, Yang J and Zhang H (2022) Exploring Immune-Related Prognostic Signatures in the Tumor Microenvironment of Colon Cancer.
Front. Genet. 13:801484.
doi: 10.3389/fgene.2022.801484

INTRODUCTION

Colon cancer is known as one of the most malignant tumors with a high mortality rate worldwide (Siegel et al., 2018). Despite the recent progress in diagnosis and therapy, the overall prognosis for colon cancer patients remains poor because effective biomarkers for prognosis prediction are lacking (Keum and Giovannucci, 2019). Therefore, it is urgent and essential to explore valuable prognostic signatures and therapeutic targets for colon cancer.

Immunotherapy takes advantage of the body's own immune system to attack cancer, which has become a powerful and promising clinical strategy for treating various tumors (Riley et al., 2019), including colon cancer(Chalabi et al., 2020; Lichtenstern et al., 2020; Ghonim et al., 2021). Immune checkpoint inhibitors (ICIs), a typical type of immunotherapy, function through inhibiting negative regulatory receptors, such as programmed cell death 1 (*PD-1*) and cytotoxic T lymphocyte antigen 4 (*CTLA4*), and thereby activates antitumor immunity (Tolba, 2020). However, only a fraction of patients were benefited from immunotherapy due to the heterogeneity and complexity of the tumor immune microenvironment (Dienstmann et al., 2017; Wang et al., 2019). Although it has been proved that IRGs were associated with the development of colon cancer (Cereda et al., 2016; Yu et al., 2019), these insights have not been applied to clinical practice. Recently, using bioinformatics and machine learning methods, various types of immune-related biomarkers have been found to be associated with the prognosis of colon cancer, such as long non-coding RNAs (Yilin Lin et al., 2020), cell infiltration (Zhou et al., 2019), and IRGs (Chen et al., 2020). However, the molecular characteristics describing the tumor immune microenvironment need to be further investigated due to their potential of prognosis and immunotherapy of colon cancer.

In this study, we constructed and validated a robust immune-related prognostic model based on TCGA-COAD cohorts and two independent GEO datasets. Additionally, we explored the relationship between the constructed prognostic model and colon cancer patients' clinical and pathological features. In addition, we analyzed the characteristics of the tumor immune microenvironment, including tumor-infiltrating cell composition, TMB, *TP53* mutation rates, and the mRNA expression levels of *PD-1/PD-L1/CLTA4*. Furthermore, the immune-related signature was also significantly associated with OS in patients with anti-PD-L1 treatment, and colon cancer patients with low-risk scores may be more sensitive to ICI therapy. These findings may provide new insights toward novel therapeutic targets for colon cancer.

MATERIALS AND METHODS

Data Acquiring

TCGA Cohorts and the List of Immune-Related Genes

The mRNA sequencing data, mutation profiling data, and clinical information were downloaded from the UCSC Xena platform (https://xenabrowser.net/datapages/). Subsequently, the samples (n = 471) were divided into normal (n = 39) and tumor groups (n = 432), and the detailed information is shown in **Supplementary Table S1**. The list of immune-related genes was acquired from the ImmPort database (https://immport.niaid.nih.gov/), with a total of 1509 genes.

GEO Cohort for External Validation

Two independent datasets (GSE39582 and GSE17536) were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The GSE39582 included 556 colon cancer samples, and GSE17536 included 177 colon cancer samples, with clinical and survival information. The detailed information is shown in **Supplementary Tables S2, S3**, respectively.

Immunotherapeutic Cohort

An immunotherapeutic cohort (IMvigor210) was obtained from a published study (Mariathasan et al., 2018), which investigated the clinical activity of the PD-L1 blockade with atezolizumab (anti-PD-L1 McAb) in urothelial cancer. The detailed clinical information and gene expression profile of the cohort were available according to the guideline on http://research-pub.gene.com/IMvigor210CoreBiologies using the IMvigor210CoreBiologies R package. We divided the samples into platinum-treated (N = 105) and non-platinum-treated datasets (N = 237) according to whether they received platinum-based chemotherapy or not, and the detailed information is shown in **Supplementary Table S4**.

Screening Immune-Related DEGs

DEGs between normal and tumor groups were screened using the *limma* R package (Ritchie et al., 2015), with the cutoff criteria set as | log2 fold change (FC)| >0.585 and adjusted *p-value* < 0.05. The immune-related DEGs were obtained by overlapping the IRGs and DEGs. In order to investigate biological pathways correlated with immune response, we performed gene ontoloy (GO) functional annotations and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis on immune-related DEGs using the *clusterProfiler* R package (Yu et al., 2012).

Construction and Validation of the Immune-Related Prognostic Model for Colon Cancer

The whole dataset (n = 432) with all tumor samples was randomly divided into training dataset (n = 216) and test dataset (n = 216) with a 1:1 ratio. As shown in the **Supplementary Table S5**, there was no significant difference among the whole dataset, the training dataset, and test dataset for most clinical-pathological factors. The training dataset was used to identify the prognostic signature and constructed a prognostic risk model. First, we identified the candidate prognostic signature using the univariable Cox proportional hazards regression model and *Survival* R package. To avoid over-fitting, all genes with *p-value* < 0.05 were involved in the subsequent least absolute shrinkage and selection operator (LASSO) analysis using the *glmnet* R package. The association between the mRNA expression level of the filtered candidate prognostic genes and patients' OS was further investigated using Kaplan–Meier analysis. Then, the multivariate Cox regression analysis was conducted to determine each independent prognostic indicator. Accordingly, the coefficient of the immune-related indicator was obtained from the multivariate Cox results. A formula for the immune-related risk score model was established to predict patient survival:

$$\text{risk}\, score = \Sigma Cox\, coefficient\, of\, gene\, \chi i \times scale\, expression\, value\, of\, gene\, \chi i.$$

To evaluate the predictive efficiency of the constructed immune-related risk score model, we plotted a receiver operating characteristic (ROC) curve to quantify the area under the curve (AUC) using the *survivalROC* R package. Also, we selected the turning point of the ROC curve with the most significant difference between true positive and false positive as the optimal cutoff risk score. Patients above the cutoff value belong to the high-risk group, while patients below it belong to the low-risk group. In addition, Kaplan–Meier curves were plotted to distinguish the two groups using the *survminer* R package.

Moreover, the test dataset and the whole dataset were used to validate the prognostic capability of the immune-related signature. Similarly, the two datasets were divided into high- and low-risk groups based on the constructed risk score model. Next, the ROC and Kaplan–Meier curves were plotted to validate the predictive accuracy of the risk score model. Then, the nomogram was constructed using the whole dataset based on the risk score model and clinical factors, including the age, gender, microsatellite status, and tumor stage. The constructed nomogram was further assessed by calibration. Additionally, the associations between the immune-related constructed risk score model and the clinical and pathological characteristics, including advanced pathological stages and TNM stages, were explored by using the Wilcoxon test. Additionally, the constructed model was further validated using GEO datasets with accession numbers GSE39582 and GSE17536.

Estimation and Comparison of Tumor-Infiltrating Immune Cell Type Fractions

The whole dataset was divided into high- and low-risk groups according to the constructed risk model, and the CIBERSORT algorithm was conducted to access the proportions of 22 types of tumor-infiltrating immune cells using the normalized gene expression matrixes and running with 1000 permutations (Newman et al., 2015). Subsequently, the comparison of immune landscape between the high- and low-risk groups was evaluated using the unpaired *t*-test. The significant differential immune cell types (*p-value* < 0.001) were further assessed for their relationship with OS using Kaplan–Meier curves.

Characteristics of Immunotherapy-Related Predictors for Colon Cancer Patients

We first calculated the TMB value and visualized the mutation profiles of the high- and low-risk groups of the whole dataset using the *maftools* R package (Mayakonda et al., 2018). The unpaired *t*-test statistically analyzed the differences of the TMB between the high- and low-risk groups. In addition, the OS between the high- and low-risk groups was calculated using the Kaplan–Meier method. Moreover, the Wilcoxon test was used to compare the mRNA levels of immune checkpoints and their ligands between the high- and low-risk groups.

Exploring the Associations Between the Microsatellite Status and the Constructed Prognostic Model

The whole dataset, after removing the samples without microsatellite status information, was used for further analysis based on the constructed prognostic model. Subsequently, the samples were divided into MSI-H and MSS/MSI-L groups according to the microsatellite status information extracted from the phenotypic data, and the Wilcoxon test was performed to analyze the difference of the level of risk score between the MSI-H and MSS/MSI-L groups. Moreover, the OS between the MSI-H group and MSI-L/MSS group was calculated using the Kaplan–Meier method.

The Role of the Immune-Related Prognostic Signature in Immunotherapy

In order to investigate the potential value of the identified immune-related signature with respect to immunotherapy, we obtained the gene expression profiles and corresponding clinical features from an immunotherapeutic cohort (Imvigor210) treated with anti-PD-L1 agent, which was divided into platinum-treated and non-platinum-treated datasets. We first validated the constructed immune-related prognostic model using the platinum-treated and non-treated datasets, respectively. Then, the complete response (CR) or partial response (PR) patients were categorized as responders and compared with non-responders, who displayed stable (SD) or progressive disease (PD), and the risk score of each patient was calculated based on the constructed risk score model. Subsequently, we statistically analyzed the distribution of the risk score between the responders and non-responders. Moreover, we further compared the tumor mutation load and neoantigen burden between high-and low-risk groups using the Wilcoxon test.

RESULTS

Identification of Immune-Related DEGs

A flow chart of the whole analysis pipeline is shown in **Figure 1**. A total of 571 DEGs (275 upregulated and 296 downregulated) were screened by comparing between tumor and normal groups (**Figure 2A**). After the intersection with 1509 IRGs, 102 immune-related DEGs were obtained (**Figure 2B**), of which 83 genes were downregulated, and 19 genes were upregulated. Detailed information is shown in **Supplementary Table S6**. Subsequently, functional and pathway enrichment

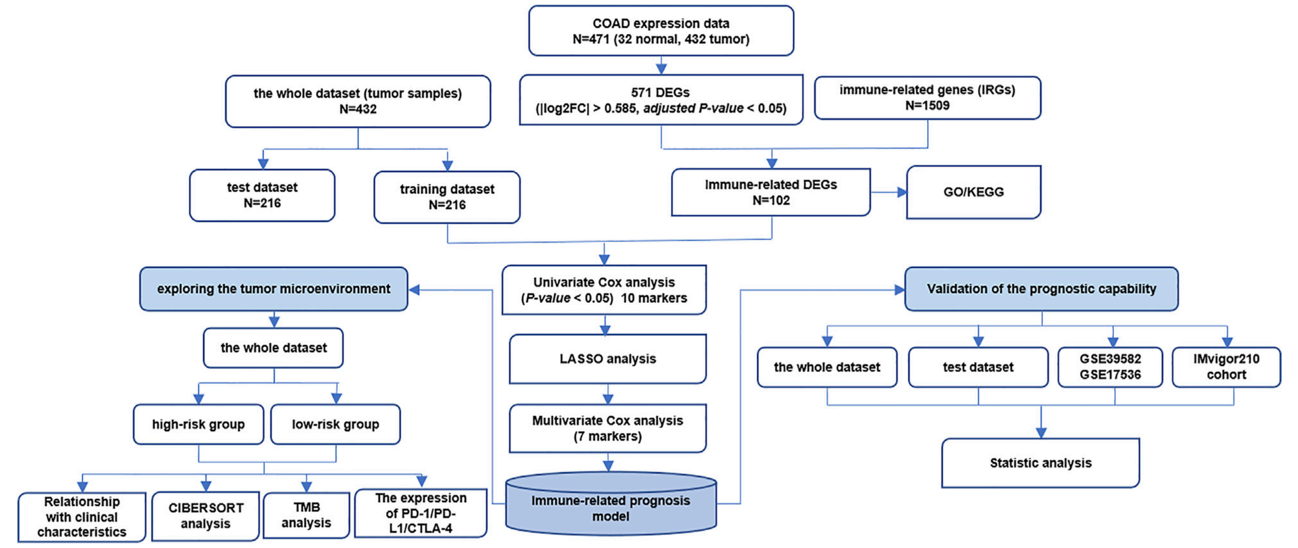


FIGURE 1 | The whole flow chart of data analysis.

analyses were performed using the *clusterProfiler* R package. KEGG analysis results indicated that the immune-related DEGs were significantly enriched in terms associated with the cytokine–cytokine receptor interaction, neuroactive ligand–receptor interaction, and IL-17 signaling pathway (**Figure 2C**), while GO related to humoral immune response was mediated by circulating immunoglobulin, humoral immune response, and immunoglobulin-mediated immune response (**Figure 2D**).

Construction of the Immune-Related Risk Score Model and the Evaluation of its Prognostic Ability

To explore the prognostic value of the immune-related DEGs, we performed the univariate Cox regression analysis. A total of 10 genes were significantly related to OS status, and 7 genes with the maximum prognostic value were further identified using LASSO regression analysis (**Supplementary Figures S1A,B**). The mRNA expression level of the seven genes was significantly associated with patients' OS (*GUCA2A*: $p = 0.013$; *CHGB*: $p = 0.05$; *SSTR2*: $p = 0.017$; *VIP*: $p = 0.0074$; *OXTR*: $p = 0.001$; *IL1A*: $p = 0.0035$; and *GRP*: $p = 0.016$), and the higher expression level of *IL1A* and *GUCA2A* was associated with a better patients' OS, while the other five genes were opposite (**Supplementary Figure S1C–I**). Then, we conducted the multivariate Cox regression analysis and established an immune-related risk score model based on the training dataset, and the hazard ratio of each gene is shown in **Figure 2E**. The colon cancer patients were divided into high- and low-risk groups according to the risk score calculated using the formula described in Materials and Methods. The scatter diagram in **Figure 2F** revealed that the OS tended to become worse with the increase of risk score, and the proportion of death in the high-risk group (the proportion of red dot and blue dot on the right side) was higher than that in the low-risk group. The heatmap in **Figure 2F** showed that the expression of *IL1A* and *GUCA2A* was low in the low-risk group and high in the high-risk group, while the trend of the other five genes was opposite. The Kaplan–Meier analysis results showed that high-risk score patients had worse OS than low score patients ($p < 0.0001$, **Figure 2G**). The prognostic accuracy of the risk score model was investigated as a continuous variable (**Figure 2H**). The AUC of the prognostic model for OS was 0.780 at 3 years, 0.801 at 4 years, and 0.766 at 5 years, indicating its excellent prediction performance.

Validation and Assessment of the Immune-Related Prognostic Signatures

To determine if the constructed risk core model is consistent in different populations, we performed an identical formula using the test dataset and the whole dataset. Consistent with the findings in the training dataset, patients categorized into the high-risk score group had worse OS than the patients in the low-risk score group ($p < 0.05$, **Supplementary Figures S2A,C**). The areas under the curves (AUCs) of the prognostic model were 0.642 for 3-year OS, 0.647 for 4-year OS, and 0.629 for 5-year OS using the test dataset, and 0.626 for 3-year OS, 0.663 for 4-year OS, and 0.661 for 5-year OS using the whole dataset (**Supplementary Figures S2B,D**). The Wilcoxon test showed that the higher risk score was associated with a higher T stage ($p = 0.00009$), N stages ($p = 0.0018$), metastasis ($p = 0.0064$), and advanced pathological stage ($p = 0.0034$) based on the whole dataset (**Figures 3A–D**).

To further validate the robustness of the prognostic signatures and improve the accuracy of the performance of the risk score model, we constructed a nomogram that integrated the immune-related risk score and clinical information, including the age, sex, microsatellite status, and tumor stage to quantitatively predict the prognosis of colon cancer patients in the whole dataset. In the

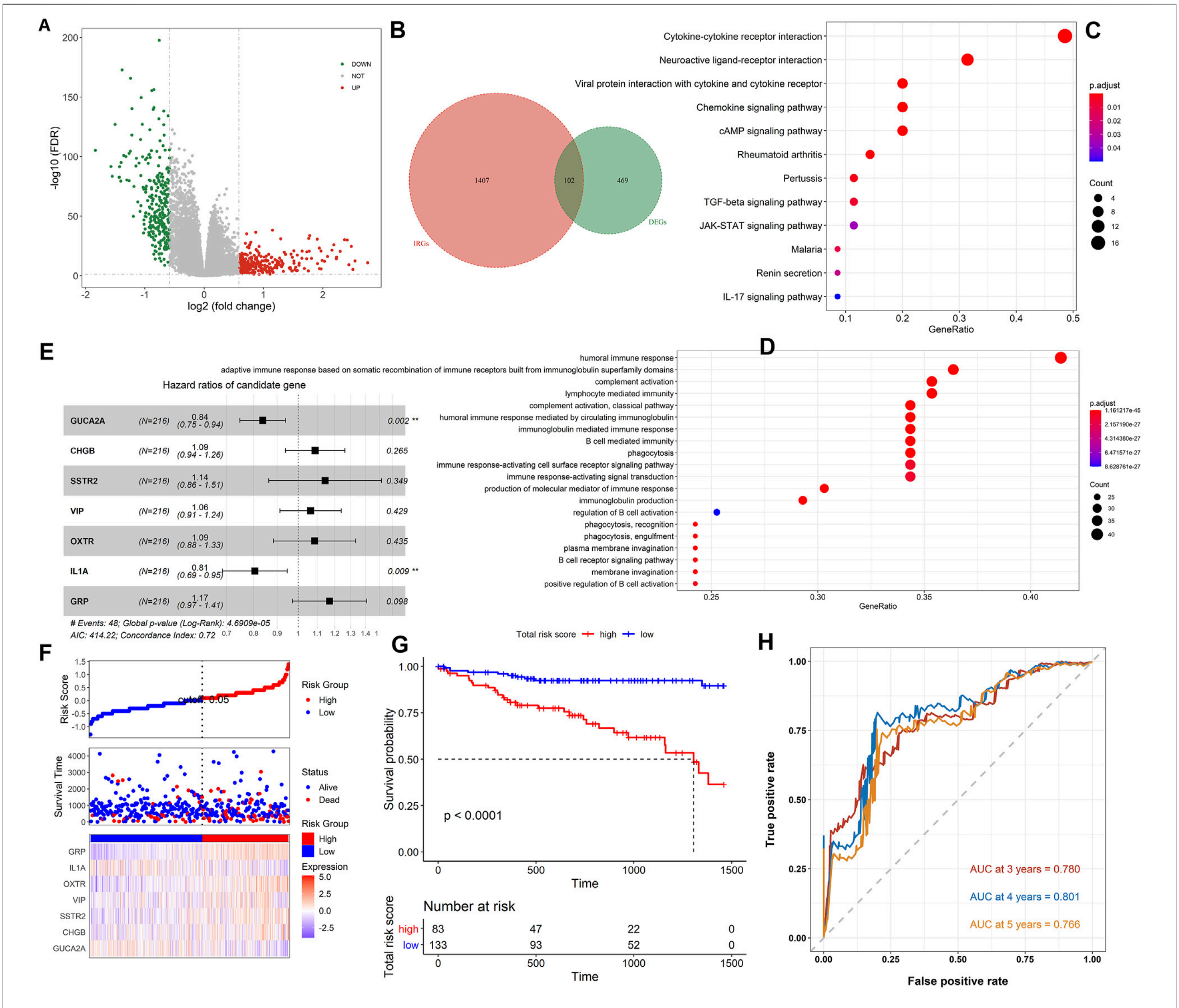


FIGURE 2 | Analysis of immune-related DEGs and construction of the immune-related prognostic model. **(A)** The volcano plot on all DEGs between the tumor and normal samples. The green dots represent downregulated genes, while the red dots represent upregulated genes. **(B)** The Venn diagram of the intersection between the DEGs and IRGs. **(C)** The KEGG pathway enrichment analysis of immune-related DEGs. **(D)** GO analysis of immune-related DEGs. **(E)** The forest plot showed the results of multivariate Cox analysis. **(F)** The distribution of the high- and low-risk score groups and their relationship with OS, and the expression pattern of seven prognostic signatures in high- and low-risk score groups. **(G)** The Kaplan–Meier curve revealed that OS in the low-risk score group was significantly higher than that in the high-risk score group. **(H)** Time-dependent ROC curve analysis of the immune-related risk score model.

nomogram, the score for each variable can be found on the point scale, so that it is easy to estimate the probability of survival at 3, 4, and 5 years by calculating the total score (**Supplementary Figure S2E**). The forest plot showed that patient's characters, including the age (>60), tumor stage (III and IV), and risk score were significantly associated with the OS (*p-value* < 0.05, **Figure 3E**). The calibration curves revealed that the predictive curves were close to the ideal curve (**Figures 3F–H**), indicating good performance. Furthermore, the predictive accuracy of this nomogram (C-index: 0.74) was higher than that of the risk score model (C-index: 0.72).

Exploring the Tumor Immune Microenvironment in Colon Cancer Patients

Based on the CIBERSORT algorithm, we estimated the proportions of 22 types of immune cells in each colon cancer patient. Then, we compared the proportions of immune cells between the low-risk group and high-risk group, and the significant differences were found in resting NK cells, M0 macrophages, M2 macrophages, CD4 memory-activated T cells, plasma cells, resting mast cells, and neutrophils. Among them, the resting NK cells and M0 macrophages were

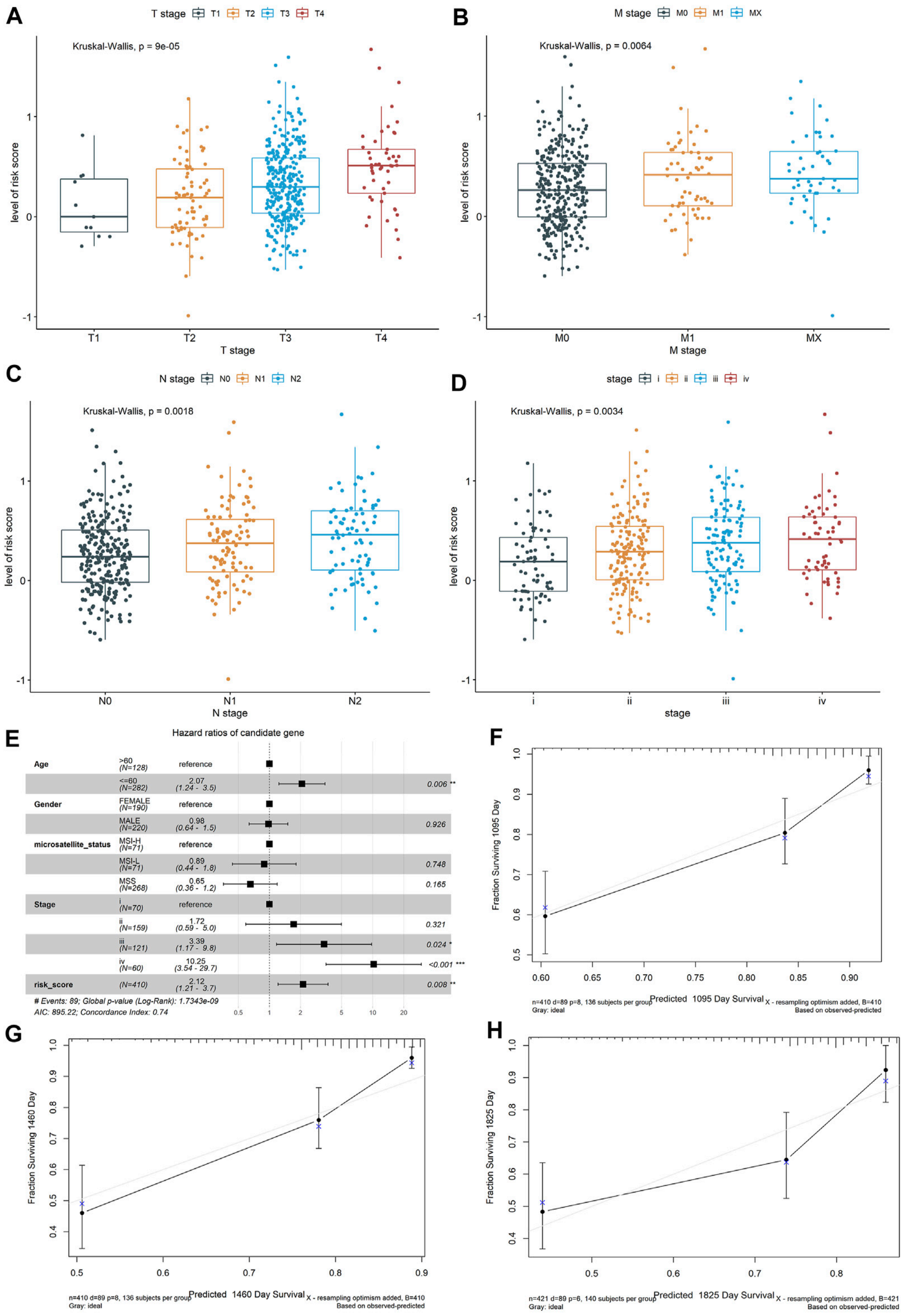


FIGURE 3 | Exploring the relationship between the risk score of the colon cancer patients and clinical and pathological characteristics, including **(A)** T stages, **(B)** M stages, **(C)** N stages, and **(D)** advanced pathological stages, based on the whole dataset. Construction and validation of a nomogram. **(E)** Forest plots showed the associations between patients' characteristics and OS. **(F–H)** The calibration plot of the nomogram to predict the probability of OS at 3, 4, and 5 years.

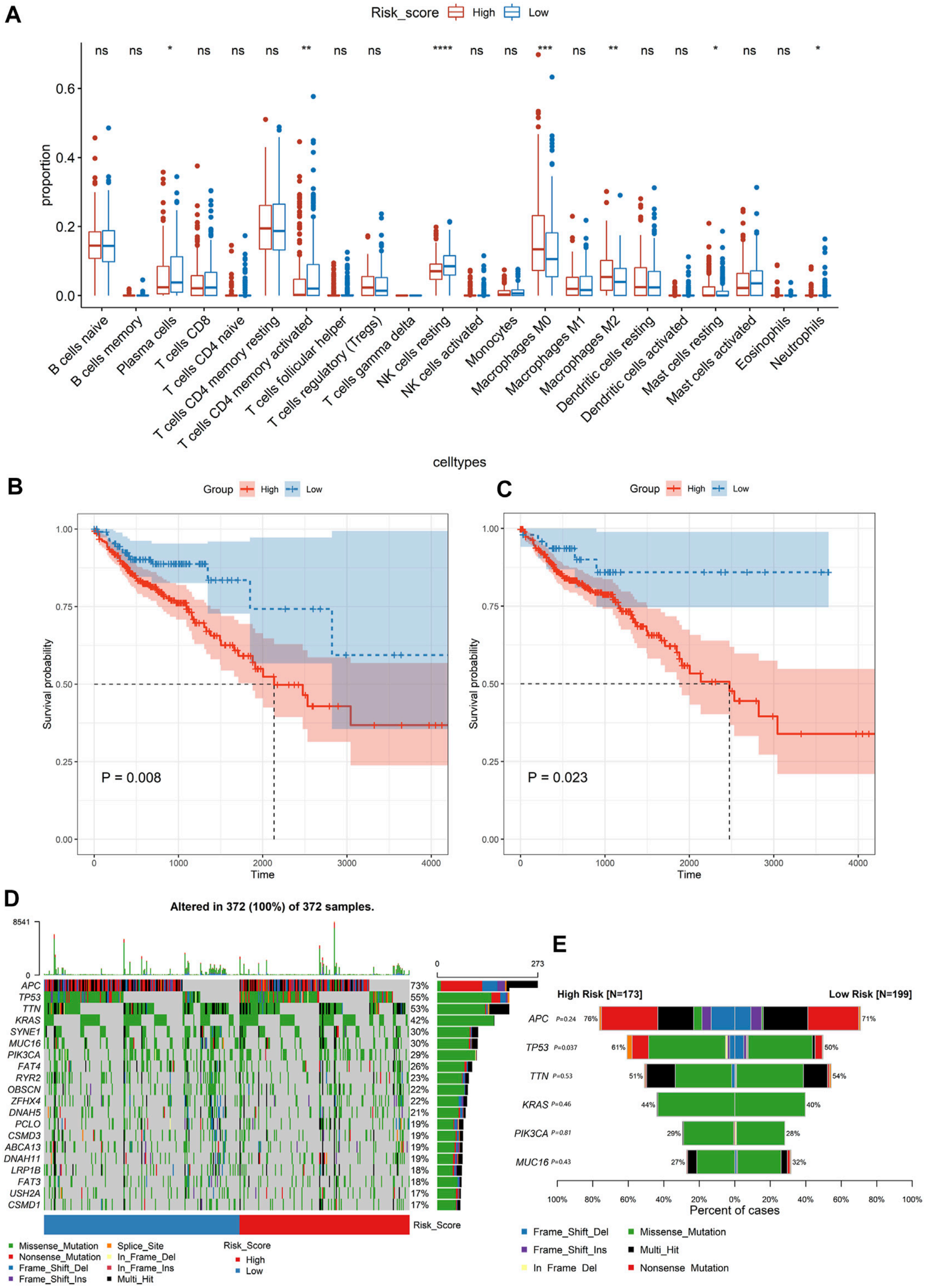


FIGURE 4 | Analysing the immune cell types and mutation profiles in high- and low-risk groups based on the whole dataset. **(A)** Comparing the difference of the proportions of immune cells between the low-risk group and high-risk group using the Wilcoxon test. The values of *P* were labeled above each boxplot with asterisks (**p* < 0.05, ***p* < 0.01, ****p* < 0.001, *****p* < 0.0001). **(B, C)** The Kaplan–Meier analysis of the relationship between the level of resting NK cells and M0 macrophages with patients' OS. **(D)** The mutation profiles of colon cancer patients in high- and low-risk groups. **(E)** Comparison of the mutation rate between the high-risk group and low-risk group.

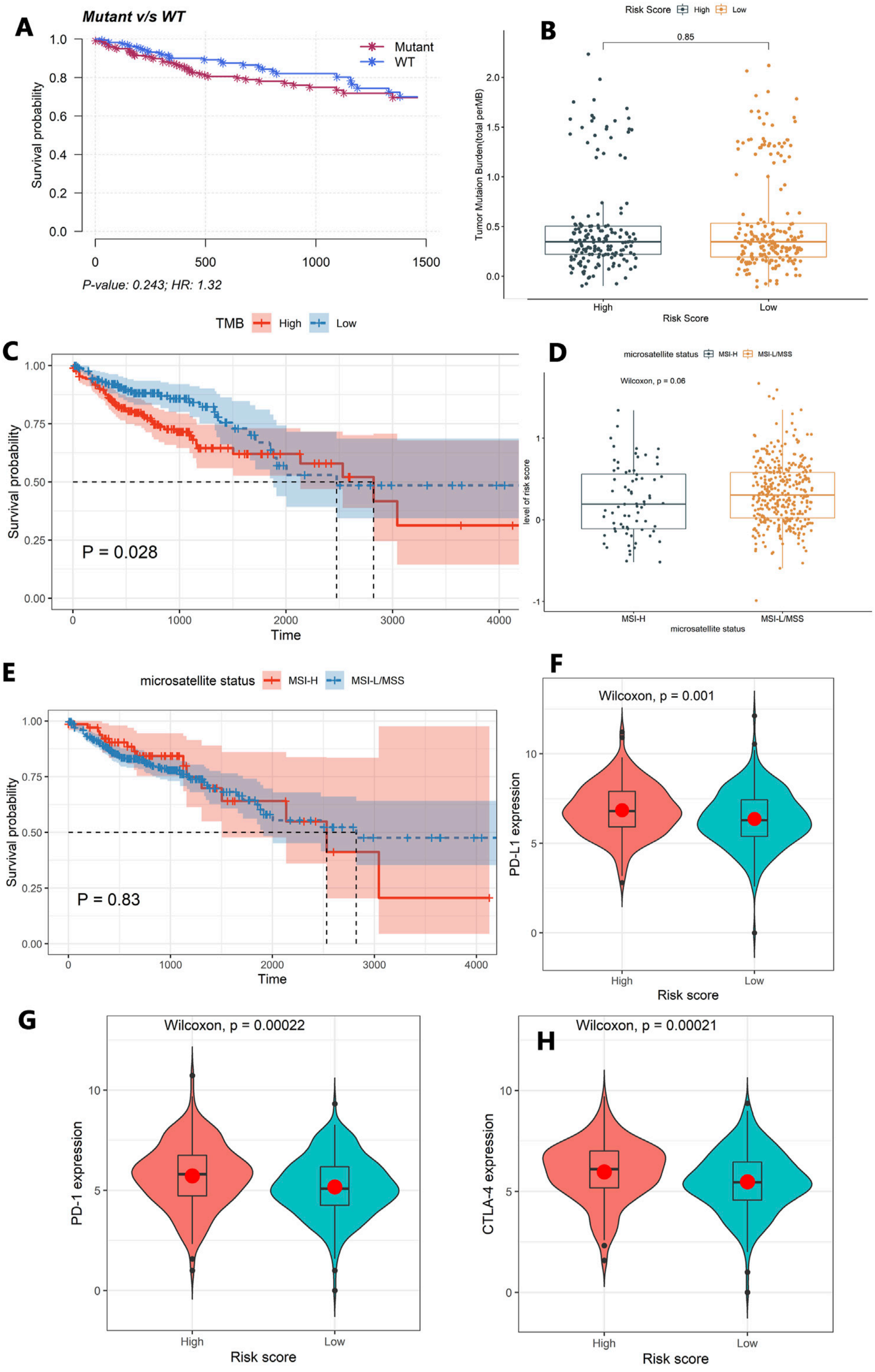


FIGURE 5 | Exploring the tumor immune microenvironment in colon cancer patients. **(A)** The association between the *TP53* status and patients' OS. **(B)** The difference of TMB between the high-risk group and the low-risk group. **(C)** The association between TMB and patients' OS. **(D)** The difference in the level of the risk score between the MSI-H and MSI-L/MSS groups. **(E)** The association between the microsatellite status and patients' OS. **(F–H)** Comparison of the expression levels of the immune checkpoints and their ligands between the high-risk score group and low-risk score group. **(F)** The expression of PD-L1, **(G)** the expression of PD-1, and **(H)** the expression of CTLA-4.

the most significant, with $p < 0.001$ (**Figure 4A**). The Kaplan–Meier curve revealed that a high level of resting NK cells or M0 macrophages was significantly related to poor OS ($p < 0.05$, **Figures 4B,C**).

The mutation profiles of each colon cancer patient were plotted using the whole dataset. As shown in **Figure 4D**, the top 20 significantly mutated genes were *APC, TP53, TTN, KRAS, SYNE1, MUC16, PIK3CA, FAT4, RYR2, OBSCN, ZFHX4, DNAH5, PCLO, CSMD3, ABCA13, DNAH11, LRP1B, FAT3, USH2A*, and *CSMD1*. Among them, the mutation rate of *TP53* was significantly different between the high-risk score group and low-risk group ($p = 0.037$, **Figure 4E**). However, the *TP53* status was not significantly related to patients' OS (**Figure 5A**). Besides, we calculated the TMB of each sample and found that there was no significant difference between the high-risk group and the low-risk group ($p = 0.85$, **Figure 5B**). However, we observed that high TMB was significantly related to poor patients' OS (**Figure 5C**). Additionally, the Wilcoxon test statistically analyzed the difference in the level of risk scores between the MSI-H and MSI-L/MSS groups, and the result showed the difference was not significant ($p = 0.06$, **Figure 5D**). As shown in **Figure 5E**, the microsatellite status cannot be used as an independent prognostic factor ($p = 0.83$).

Moreover, the Wilcoxon test was used to compare the expression levels of the immune checkpoints and their ligands between the high-risk score group and low-risk score group. The high-risk score group had a high expression level of *PD-L1* ($p = 0.001$), *PD-1*($p = 0.00022$), and *CTLA-4* ($p = 0.00021$, **Figures 5F–H**).

Validation of the Prognostic Signature Using GEO Datasets

In order to investigate the applicability in multiple cohorts based on different platforms, we further verified the constructed risk score model using two independent external GEO datasets. We found that patients with high-risk scores demonstrated worse prognosis than those with low-risk scores (GSE39582: $p = 0.0023$, **Supplementary Figure S3**; GSE17536: $p = 0.0008$, **Supplementary Figure S4**). The AUCs of the prognostic model were 0.577 for 3-year OS, 0.569 for 4-year OS, and 0.568 for 5-year OS using the GSE39582 dataset (**Supplementary Figure S3A**), and 0.578 for 3-year OS, 0.699 for 4-year OS, and 0.657 for 5-year OS using GSE17536 (**Supplementary Figure S4A**). The Wilcoxon test showed that a higher risk score was associated with a higher T stage ($p = 0.0041$), metastasis ($p = 0.04$), N stages ($p = 0.037$), and advanced pathological stages ($p = 0.033$) using GSE39582 (**Supplementary Figures S3C–F**). It was also found that a higher risk score was associated with a higher advanced pathological stage ($p = 0.029$) in GSE17536 (**Supplementary Figure S4C**), without obtaining the TMN stage data.

The Prognostic Signature in the Role of ICI Treatment

In the immunotherapeutic cohort, patients with a low-risk score exhibited a significantly prolonged survival rate (non-platinum-treated dataset: $p = 0.0014$, **Supplementary Figure S5A**; platinum-treated dataset: $p = 0.033$, **Supplementary Figure S6A**). Patients without platinum treatment indicated marked clinical benefits from immunotherapy in the low-risk score group compared to those with a high-risk score ($p = 0.0027$, **Supplementary Figure S5B**), but not significantly in patients with platinum treatment ($p = 0.44$, **Supplementary Figure S6B**). Further analysis revealed that a higher tumor mutation load in patients with platinum treatment was significantly associated with a low-risk score ($p = 0.039$, **Supplementary Figure S6C**), but not in patients without platinum treatment ($p = 0.21$, **Supplementary Figure S5C**). Interestingly, higher neoantigen burden in patients without platinum treatment was significantly associated with a low-risk score ($p = 0.025$, **Supplementary Figure S5D**), but not in patients with platinum treatment ($p = 0.5$, **Supplementary Figure S6D**).

DISCUSSION

The immune cells within the tumor microenvironment (TME) function play a key role in tumorigenesis (Lei et al., 2020). Growing evidence has revealed the therapeutic potential of ICIs in colon cancer (Kimura et al., 2020; Wang et al., 2020). However, the limited knowledge on the characteristics of the TME, to some extent, hindered the development of the application of immunotherapy. In recent years, many efforts have been made to identify immune-related biomarkers for the diagnosis and prognosis of colon cancer (Zhou et al., 2019; Laghi et al., 2020; Li et al., 2020). However, more reliable biomarkers still need to be explored to maximize the application of immunotherapy.

In this study, we developed a prognostic risk score model based on seven IRGs, named *GUCA2A, CHGB, SSTR2, VIP, OXTR, IL1A*, and *GRP*, which has been verified in multiple cohorts across different platforms. Among them, *GUCA2A, VIP*, and *OXTR* have been demonstrated to be significantly associated with the prognosis of colon cancer (Zhang et al., 2019; Houxi Xu et al., 2020; Kang Lin et al., 2020; Zhang et al., 2020). A previous study reported that guanylyl cyclase C (*GUCY2C*) could act as a tumor suppressor and play an important role in orchestrating intestinal homeostatic mechanisms, which could be used as a therapeutic target for colon cancer patients (such as the FDA-approved oral *GUCY2C* ligand linaclotide, Linzess™) (Pattison et al., 2016). *GUCA2A* may be considered as a potential marker for the prognosis and therapeutic target in colon cancer by binding and activating *GUCY2C*. As a precursor of regulatory peptide, the relationship between *CHGB* and tumor is not clear. However, *CHGB* was suggested to be an immune-related signature for low-grade glioma (Liu et al., 2021) and head and neck squamous cell carcinoma (Zhang et al., 2021). Previous studies also experimentally demonstrated that an abnormal expression of *CHGB* was associated with aggressive VHL-associated pancreatic neuroendocrine tumors (validated by immunohistochemistry) (Weisbrod et al., 2013), pancreatic cancer (validated by qPCR) (Jia-Sheng Xu et al., 2020), and small cell lung cancer (validated by immunoblotting and qPCR) (Moss et al., 2009). *SSTR2*, as a G protein-coupled cell surface receptor, can be activated by extracellular ligands, which leads to the inhibition of cell proliferation (Lechner et al., 2021). Precious studies demonstrated that *SSTR2* might serve as a molecular target in the diagnosis and treatment of thyroid cancer (Thakur et al., 2021), small intestinal neuroendocrine tumor (Elf et al., 2021), and neuroendocrine tumors (Si et al., 2021). *VIP* can provide protection from apoptosis in tumorigenesis (Sastry et al., 2017). OXTR and its ligand oxytocin (OXT) are present in the gastrointestinal system and involved in tumorigenesis (Ma et al., 2019). *IL1A* was involved in various immune responses, inflammatory processes, and hematopoiesis, which might be associated with colon tumorigenesis (Yoshikawa et al., 2017). To our knowledge, the relationship between *GRP* and tumorigenesis has not been reported.

Furthermore, we systematically explored the characteristics of the tumor immune microenvironment. The results revealed that the tumor-infiltrating resting NK cells or M0 Macrophages, *TP53* mutation rates, and *TMB* could be independent prognostic signatures for colon cancer. Additionally, we observed that the expression levels of checkpoint genes (*PD-L1, PD-1*, and *CTLA-4*) were higher in high-risk score patients, which may suggest that our immune-related risk score model was capable of providing support for immunotherapy. More importantly, the immune-related signature was also significantly associated with OS in patients with anti-PD-L1 treatment. We speculated that patients with a low-risk score might be more sensitive to ICI therapy based on the result of **Supplementary Figures S5, S6**.

In addition, we compared the performance of our constructed immune-related prognostic model with the published prognostic model of colon cancer based on the cohorts TCGA-COAD, GSE39582, and GSE17536, which is summarized in **Supplementary Table S7**. Our constructed prognostic model was relatively and effectively validated in more internal and external cohorts, including an immunotherapeutic cohort.

CONCLUSION

In summary, a robust immune-related prognostic model was constructed, and the characteristics of the tumor immune microenvironment were explored, which may be helpful for the prognosis and immunotherapy of colon cancer patients.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found in the following: the expression profiles and corresponding clinical information of TCGA-COAD were downloaded from https://xenabrowser.net/datapages/. GSE39582 and GSE17536 were downloaded from the NCBI-GEO database (https://www.ncbi.nlm.nih.gov/geo/); and the immunotherapeutic cohort (IMvigor210) was available according to the guideline on http://research-pub.gene.com/IMvigor210CoreBiologies using the IMvigor210CoreBiologies R package.

AUTHOR CONTRIBUTIONS

LC and EC were involved in the study concept, design, and drafting of the manuscript. HZ and YB put forward some kind suggestions. TL and JY helped in analyzing and interpreting the data.

FUNDING

This work was supported by a grant from the Key Science and Technology Program of the Shaanxi Province (2019ZDLSF02-05) and the National Natural Scientific Foundation of China (81974378 and 82003115).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.801484/full#supplementary-material

REFERENCES

Cereda, M., Gambardella, G., Benedetti, L., Iannelli, F., Patel, D., Basso, G., et al. (2016). Patients with Genetically Heterogeneous Synchronous Colorectal Cancer Carry Rare Damaging Germline Mutations in Immune-Related Genes. *Nat. Commun.* 7, 12072. doi:10.1038/ncomms12072

Chalabi, M., Fanchi, L. F., Dijkstra, K. K., Van den Berg, J. G., Aalbers, A. G., Sikorska, K., et al. (2020). Neoadjuvant Immunotherapy Leads to Pathological Responses in MMR-Proficient and MMR-Deficient Early-Stage colon Cancers. *Nat. Med.* 26 (4), 566–576. doi:10.1038/s41591-020-0805-8

Chen, H., Luo, J., and Guo, J. (2020). Development and Validation of a Five-Immune Gene Prognostic Risk Model in colon Cancer. *BMC Cancer* 20 (1), 395. doi:10.1186/s12885-020-06799-0

Dienstmann, R., Vermeulen, L., Guinney, J., Kopetz, S., Tejpar, S., and Tabernero, J. (2017). Consensus Molecular Subtypes and the Evolution of Precision Medicine in Colorectal Cancer. *Nat. Rev. Cancer* 17 (2), 79–92. doi:10.1038/nrc.2016.126

Elf, A.-K., Johanson, V., Marin, I., Bergström, A., Nilsson, O., Svensson, J., et al. (2021). Evaluation of SSTR2 Expression in SI-NETs and Relation to Overall Survival after PRRT. *Cancers* 13 (9), 2035. doi:10.3390/cancers13092035

Ghonim, M. A., Ibba, S. V., Tarhuni, A. F., Errami, Y., Luu, H. H., Dean, M. J., et al. (2021). Targeting PARP-1 with Metronomic Therapy Modulates MDSC Suppressive Function and Enhances Anti-PD-1 Immunotherapy in colon Cancer. *J. Immunother. Cancer* 9 (1), e001643. doi:10.1136/jitc-2020-001643

Keum, N., and Giovannucci, E. (2019). Global burden of Colorectal Cancer: Emerging Trends, Risk Factors and Prevention Strategies. *Nat. Rev. Gastroenterol. Hepatol.* 16 (12), 713–732. doi:10.1038/s41575-019-0189-8

Kimura, Y., Sumiyoshi, M., Kiyoi, T., and Baba, K. (2020). Dihydroxystilbenes Prevent Azoxymethane/dextran Sulfate Sodium-Induced colon Cancer by Inhibiting colon Cytokines, a Chemokine, and Programmed Cell Death-1 in C57BL/6J Mice. *Eur. J. Pharmacol.* 886, 173445. doi:10.1016/j.ejphar.2020.173445

Laghi, L., Negri, F., Gaiani, F., Cavalleri, T., Grizzi, F., de' Angelis, G. L., et al. (2020). Prognostic and Predictive Cross-Roads of Microsatellite Instability and Immune Response to Colon Cancer. *Int. J. Mol. Sci.* 21 (24), 9680. doi:10.3390/ijms21249680

Lechner, M., Schartinger, V. H., Steele, C. D., Nei, W. L., Ooft, M. L., Schreiber, L.-M., et al. (2021). Somatostatin Receptor 2 Expression in Nasopharyngeal Cancer Is Induced by Epstein Barr Virus Infection: Impact on Prognosis, Imaging and Therapy. *Nat. Commun.* 12 (1), 117. doi:10.1038/s41467-020-20308-8

Lei, X., Lei, Y., Li, J.-K., Du, W.-X., Li, R.-G., Yang, J., et al. (2020). Immune Cells within the Tumor Microenvironment: Biological Functions and Roles in Cancer Immunotherapy. *Cancer Lett.* 470, 126–133. doi:10.1016/j.canlet.2019.11.009

Li, X., Wen, D., Li, X., Yao, C., Chong, W., and Chen, H. (2020). Identification of an Immune Signature Predicting Prognosis Risk and Lymphocyte Infiltration in Colon Cancer. *Front. Immunol.* 11, 1678. doi:10.3389/fimmu.2020.01678

Lichtenstern, C. R., Ngu, R. K., Shalapour, S., and Karin, M. (2020). Immunotherapy, Inflammation and Colorectal Cancer. *Cells* 9 (3), 618. doi:10.3390/cells9030618

Lin, K., Huang, J., Luo, H., Luo, C., Zhu, X., Bu, F., et al. (2020). Development of a Prognostic index and Screening of Potential Biomarkers Based on Immunogenomic Landscape Analysis of Colorectal Cancer. *Aging* 12 (7), 5832–5857. doi:10.18632/aging.102979

Lin, Y., Pan, X., Chen, Z., Lin, S., and Chen, S. (2020). Identification of an Immune-Related Nine-lncRNA Signature Predictive of Overall Survival in Colon Cancer. *Front. Genet.* 11, 318. doi:10.3389/fgene.2020.00318

Liu, W., Zou, J., Ren, R., Liu, J., Zhang, G., and Wang, M. (2021). A Novel 10-Gene Signature Predicts Poor Prognosis in Low Grade Glioma. *Technol. Cancer Res. Treat.* 20, 153303382199208. doi:10.1177/1533033821992084

Ma, M., Li, L., Chen, H., and Feng, Y. (2019). Oxytocin Inhibition of Metastatic Colorectal Cancer by Suppressing the Expression of Fibroblast Activation Protein-α. *Front. Neurosci.* 13, 1317. doi:10.3389/fnins.2019.01317

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., et al. (2018). TGFβ Attenuates Tumour Response to PD-L1 Blockade by Contributing to Exclusion of T Cells. *Nature* 554 (7693), 544–548. doi:10.1038/nature25501

Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H. P. (2018). Maftools: Efficient and Comprehensive Analysis of Somatic Variants in Cancer. *Genome Res.* 28 (11), 1747–1756. doi:10.1101/gr.239244.118

Moss, A. C., Jacobson, G. M., Walker, L. E., Blake, N. W., Marshall, E., and Coulson, J. M. (2009). SCG3 Transcript in Peripheral Blood Is a Prognostic Biomarker for REST-Deficient Small Cell Lung Cancer. *Clin. Cancer Res.* 15 (1), 274–283. doi:10.1158/1078-0432.CCR-08-1163

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Pattison, A. M., Merlino, D. J., Blomain, E. S., and Waldman, S. A. (2016). Guanylyl Cyclase C Signaling axis and colon Cancer Prevention. *World J. Gastroenterol.* 22 (36), 8070–8077. doi:10.3748/wjg.v22.i36.8070

Riley, R. S., June, C. H., Langer, R., and Mitchell, M. J. (2019). Delivery Technologies for Cancer Immunotherapy. *Nat. Rev. Drug Discov.* 18 (3), 175–196. doi:10.1038/s41573-018-0006-z

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Sastry, K. S., Chouchane, A. I., Wang, E., Kulik, G., Marincola, F. M., and Chouchane, L. (2017). Cytoprotective Effect of Neuropeptides on Cancer Stem Cells: Vasoactive Intestinal Peptide-Induced Antiapoptotic Signaling. *Cell Death Dis* 8 (6), e2844. doi:10.1038/cddis.2017.226

Si, Y., Kim, S., Ou, J., Lu, Y., Ernst, P., Chen, K., et al. (2021). Anti-SSTR2 Antibody-Drug Conjugate for Neuroendocrine Tumor Therapy. *Cancer Gene Ther.* 28 (7-8), 799–812. doi:10.1038/s41417-020-0196-5

Siegel, R. L., Miller, K. D., and Jemal, A. (2018). Cancer Statistics, 2018. *CA: A Cancer J. Clinicians* 68 (1), 7–30. doi:10.3322/caac.21442

Thakur, S., Daley, B., Millo, C., Cochran, C., Jacobson, O., Lu, H., et al. (2021). 177Lu-DOTA-EB-TATE, a Radiolabeled Analogue of Somatostatin Receptor Type 2, for the Imaging and Treatment of Thyroid Cancer. *Clin. Cancer Res.* 27 (5), 1399–1409. doi:10.1158/1078-0432.CCR-20-3453

Tolba, M. F. (2020). Revolutionizing the Landscape of Colorectal Cancer Treatment: The Potential Role of Immune Checkpoint Inhibitors. *Int. J. Cancer* 147 (11), 2996–3006. doi:10.1002/ijc.33056

Wang, H., Wang, X., Xu, L., Zhang, J., and Cao, H. (2019). A Molecular Sub-cluster of colon Cancer Cells with Low VDR Expression Is Sensitive to Chemotherapy, BRAF Inhibitors and PI3K-mTOR Inhibitors Treatment. *Aging* 11 (19), 8587–8603. doi:10.18632/aging.102349

Wang, Y., Bhave, M. S., Yagita, H., and Cardell, S. L. (2020). Natural Killer T-Cell Agonist α-Galactosylceramide and PD-1 Blockade Synergize to Reduce Tumor Development in a Preclinical Model of Colon Cancer. *Front. Immunol.* 11, 581301. doi:10.3389/fimmu.2020.581301

Weisbrod, A. B., Zhang, L., Jain, M., Barak, S., Quezado, M. M., and Kebebew, E. (2013). Altered PTEN, ATRX, CHGA, CHGB, and TP53 Expression Are Associated with Aggressive VHL-Associated Pancreatic Neuroendocrine Tumors. *Horm. Canc* 4 (3), 165–175. doi:10.1007/s12672-013-0134-1

Xu, H., Ma, Y., Zhang, J., Gu, J., Jing, X., Lu, S., et al. (2020). Identification and Verification of Core Genes in Colorectal Cancer. *Biomed. Res. Int.* 2020, 1–13. doi:10.1155/2020/8082697

Xu, J. S., Liao, K. L., Wang, X., He, J., and Wang, X. Z. (2020). Combining Bioinformatics Techniques to Explore the Molecular Mechanisms Involved in Pancreatic Cancer Metastasis and Prognosis. *J. Cel. Mol. Med.* 24 (24), 14128–14138. doi:10.1111/jcmm.16023

Yoshikawa, T., Wu, J., Otsuka, M., Kishikawa, T., Suzuki, N., Takata, A., et al. (2017). Repression of MicroRNA Function Mediates Inflammation-Associated Colon Tumorigenesis. *Gastroenterology* 152 (3), 631–643. doi:10.1053/j.gastro.2016.10.043

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16 (5), 284–287. doi:10.1089/omi.2011.0118

Yu, G., Wu, Y., Wang, W., Xu, J., Lv, X., Cao, X., et al. (2019). Low-dose Decitabine Enhances the Effect of PD-1 Blockade in Colorectal Cancer with Microsatellite Stability by Re-modulating the Tumor Microenvironment. *Cell Mol Immunol* 16 (4), 401–409. doi:10.1038/s41423-018-0026-y

Zhang, H., Du, Y., Wang, Z., Lou, R., Wu, J., and Feng, J. (2019). Integrated Analysis of Oncogenic Networks in Colorectal Cancer Identifies GUCA2A as a Molecular Marker. *Biochem. Res. Int.* 2019, 1–13. doi:10.1155/2019/6469420

Zhang, H., Qin, C., Gan, H., Guo, X., and Zhang, L. (2020). Construction of an Immunogenomic Risk Score for Prognostication in Colon Cancer. *Front. Genet.* 11, 499. doi:10.3389/fgene.2020.00499

Zhang, Y., Chen, P., Zhou, Q., Wang, H., Hua, Q., Wang, J., et al. (2021). A Novel Immune-Related Prognostic Signature in Head and Neck Squamous Cell Carcinoma. *Front. Genet.* 12, 570336. doi:10.3389/fgene.2021.570336

Zhou, R., Zhang, J., Zeng, D., Sun, H., Rong, X., Shi, M., et al. (2019). Immune Cell Infiltration as a Biomarker for the Diagnosis and Prognosis of Stage I-III colon Cancer. *Cancer Immunol. Immunother.* 68 (3), 433–442. doi:10.1007/s00262-018-2289-7

Conflict of Interest: HZ and YB were employed by the company Shenzhen Nucleus Gene Technology Co., Ltd.

The remaining 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.

frontiers
in Genetics

ORIGINAL RESEARCH
published: 21 March 2022
doi: 10.3389/fgene.2022.788670

Characterization of Pyroptosis-Related Subtypes *via* RNA-Seq and ScRNA-Seq to Predict Chemo-Immunotherapy Response in Triple-Negative Breast Cancer

Chenlu Li[1†], Jingjing Pan[2†], Yinyan Jiang[3], Yanzhi Wu[4], Zhenlin Jin[3]* and Xupeng Chen[1]*

[1]Department of Gastroenterology, Affiliated Yueqing Hospital, Wenzhou Medical University, Wenzhou, China, [2]Department of Laboratory Medicine, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, [3]Department of Hematopathology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, [4]Department of Rheumatology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

Edited by:
Luis Zapata,
Institute of Cancer Research (ICR), United Kingdom

Reviewed by:
Feng Gao,
The Sixth Affiliated Hospital of Sun Yat-sen University, China
Yuchen Yang,
Sun Yat-sen University, China

***Correspondence:**
Zhenlin Jin
jinzhenlin163@163.com
Xupeng Chen
13506551205@qq.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *03 October 2021*
Accepted: *15 February 2022*
Published: *21 March 2022*

Citation:
Li C, Pan J, Jiang Y, Wu Y, Jin Z and Chen X (2022) Characterization of Pyroptosis-Related Subtypes via RNA-Seq and ScRNA-Seq to Predict Chemo-Immunotherapy Response in Triple-Negative Breast Cancer. Front. Genet. 13:788670. doi: 10.3389/fgene.2022.788670

Triple-negative breast cancer (TNBC) is associated with poor prognosis and invalid therapeutical response to immunotherapy due to biological heterogeneity. There is an urgent need to screen for reliable indices, especially immunotherapy-associated biomarkers that can predict patient outcomes. Pyroptosis, as an inflammation-induced type of programmed cell death, is shown to create a tumor-suppressive environment and improve the chemotherapeutic response in multiple tumors. However, the specific therapeutic effect of pyroptosis in TNBC remains unclear. In this study, we present a consensus clustering by pyroptosis-related signatures of 119 patients with TNBC into two subtypes (clusterA and clusterB) with distinct immunological and prognostic characteristics. First, clusterB, associated with better outcomes, was characterized by a significantly higher pyroptosis-related signature expression, tumor microenvironment prognostic score, and upregulation of immunotherapy checkpoints. A total of 262 differentially expressed genes between the subtypes were further identified and the Ps-score was built using LASSO and COX regression analyses. The external GEO data set demonstrated that cohorts with low Ps-scores consistently had higher expression of pyroptosis-related signatures, immunocyte infiltration levels, and better prognosis. In addition, external immunotherapy and chemotherapy cohorts validated that patients with lower Ps-scores exhibited significant therapeutic response and clinical benefit. Combined with other clinical characteristics, we successfully constructed a nomogram to effectively predict the survival rate of patients with TNBC. Finally, using the scRNA-seq data sets, we validated the landscape of cellular subtypes of TNBC and successfully constructed an miRNA-Ps-score gene interaction network. These findings indicated that the systematic assessment of tumor pyroptosis and identification of Ps-scores has potential clinical implications and facilitates tailoring optimal immunotherapeutic strategies for TNBC.

Keywords: pyroptosis, triple-negative breast cancer, molecular subtype, tumor microenvironment, prognosis, single-cell RNA sequencing

INTRODUCTION

As the most common tumor with high morbidity and mortality in women, breast cancer (BC) has a poor prognosis and exacerbates a critical social burden worldwide (Global Burden of Disease Cancer Collaboration et al., 2017). Triple-negative breast cancer (TNBC), accounting for 10%–17% among all BCs, is a special subtype characterized by negative human epidermal growth factor receptor 2 (HER2), progesterone receptor (PR), and estrogen receptor (ER) (Lin et al., 2012). Due to the absence of the corresponding receptors, patients with TNBC fail to benefit from endocrine targeted therapy and HER2-targeted agents; hence, chemotherapy and surgery remain the most common treatment for patients with TNBC (Bergin and Loi, 2019). Recently, the rapid rise of immunotherapy with the combination of cisplatin or other platinum drugs, including anti-programmed cell death (PD)-1 and PD-ligand 1 (PD-L1) agents, has brought a new therapeutic landscape for patients with TNBC who did not benefit from conventional chemotherapy, radiation, or surgery (Nolan et al., 2017). However, in clinical practice, some patients with TNBC are still lacking an effective therapeutical response to immunotherapy due to genetic and biological heterogeneity (Vikas et al., 2018). Therefore, it is crucial to identify novel subtypes and screen reliable biomarkers (especially immunotherapy-related biomarkers) that can predict outcomes of patients with TNBC.

In the clinical setting, the TNM stage system is acknowledged as the most frequently used tool to predict the prognosis of patients with TNBC, which majorly depends on the inherent anatomical abnormity, including tumor size, lymph node situation, and distant metastatic status (Park et al., 2019). However, the occurrence of biological and tumor genetic heterogeneity makes it challenging for the TNM system to predict disease progression and prognosis (Park et al., 2019). Pyroptosis is a form of pro-inflammatory programmed cell death (PCD) that cleaves the gasdermin D (GSDMD) protein by classical or nonclassical pathways and triggers the production and release of cytokines (including inactive cytokines like IL-18 and IL-1β) to induce a strong inflammatory response (Yang et al., 2016). Pyroptosis is reported to create a tumor-suppressive environment by releasing inflammatory factors; therefore, inducing pyroptosis in tumors via chemotherapeutic drugs could produce antitumor effects (Shi et al., 2015). *In vitro*, Nathalia et al. demonstrate that omega-3 fatty acids can induce pyroptosis in TNBC cells via inducing the active CASP1 increase, further leading to the cleavage of GSDMD, formation of membrane pores, and the release of IL-1β (Pizato et al., 2018). However, the exact contribution of pyroptosis on the therapeutical response of immunotherapies and its role in the prognosis of TNBC remains unclear.

The classification of patients with TNBC based on transcriptome profiles via next-generation sequencing is considered a novel method to quickly indicate biological characteristics and help screen for the most appropriate treatment strategies (He et al., 2018). Besides conventional expression profiles, various biological signatures are also applied to identify novel molecular subtypes for the prognosis of TNBC, such as autophagy-related signatures (Kim et al., 2012), N6-methyladenosine (Wu et al., 2021), immune cell infiltration (Harano et al., 2018), etc. In this study, we aimed to build a novel scoring model (called Ps-scores) based on pyroptosis-related signatures to identify clustering subtypes of TNBC and correlate the characteristics of each subtype with prognosis, immunotherapy, and immune cell infiltration in patients with TNBC. Combining the Ps-scores and other classical clinical features, the predicted model was established to improve prognostic risk stratification and facilitate the decision making of treatments for patients with TNBC. Moreover, using single-cell RNA sequencing (scRNA-seq) technology, we successfully validated the potential cellular subtypes of TNBC and expounded the predominant expression characteristics of Ps-score-related genes in each cluster. Finally, the targeted miRNAs were predicted by combining multiple databases with differentially expressed miRNAs (DEmiRNAs), whereas the miRNA-Ps-score signature interaction network was further constructed to visualize the potential regulatory relationship. These results imply the potential links among the pyroptosis-related scores, immune microenvironment, prognosis, and response to immunotherapy for patients with TNBC. Our findings provide new insight into the prognostic signatures of TNBC and will help develop promising strategies for TNBC immunotherapy.

MATERIALS AND METHODS

TNBC Data set Preparation and Preprocessing

Transcriptome profiling data (FPKM value) of 1217 BC samples with their corresponding clinical data were downloaded from The Cancer Genome Atlas (TCGA) data sets (https://portal.gdc.cancer.gov/). Through screening the "negative HER2, PR, and ER" status based on clinical data, we finally identified 119 patients with TNBC and comprehensive clinical information. Other microarray data sets of 819 patients with TNBC and prognostic information were also downloaded from the Gene Expression Omnibus (GEO) data sets (https://www.ncbi.nlm.nih.gov/geo/), including 120 ER-negative BC in GSE16446, 198 HER2-negative BC in GSE25065, 310 HER2-negative BC in GSE25055, 107 TNBC in GSE58812, and 84 TNBC in GSE157284. In addition, corresponding miRNA sequencing data sets and mutation files were also obtained from the TCGA-BRCA to investigate the miRNA regulatory mechanism, and transcriptome profiling of 179 normal breast tissues was obtained from the Genotype-Tissue Expression (GTEx) database as normal controls (Carithers and Moore, 2015). The "ComBat" algorithm of the "sva" package was further applied to remove the nonbiological technical biases due to batch effects between different data sets (Leek et al., 2012). To remove the false positives caused by batch effects, we selected several stable internal reference genes (*HPRT1*, *PPIA*, *RPS13*, *TBP*, *GAPDH*, and *HMBS*) to perform the PCA analysis, which are reported as valid reference genes for human BC cell lines by Liu et al. (2015). Moreover, the IMvigor210 data sets

(Hoffman-Censits et al., 2016), including 316 metastatic urothelial carcinomas (mUCs) with immunotherapy, were applied to investigate the therapeutic reaction, and the scRNA-seq data of 1534 cells from six patients with TNBC (GSE 118389) were used to validate molecular subtypes for TNBC. Detailed information on these data sets is listed in **Supplementary Table S1**.

Identification of Pyroptosis-Related Signatures

According to previous studies, the caspase family, especially caspase1/4/5/8 (CASP1/4/5/8) was reported to specifically cleave GSDMD to further activate pyroptosis (Shi et al., 2015; Orning et al., 2018). In addition, Zhang et al. (2020), also found that CASP3 and granzyme B (GZMB) could convert cell apoptosis into pyroptosis through cleaving gasdermin E (GSDME). Granzyme A (GZMA) was also considered to be essential in inducing cell pyroptosis by cleaving gasdermin B (GSDMB) (Zhou et al., 2020). Moreover, inflammasome-associated families, such as absent in melanoma 2 (AIM2) and nucleotide-binding domain and leucine-rich repeat receptor (NLR), are demonstrated to induce the pyroptosis process through activating CASP1 and the release of IL1β and IL18 (Man and Kanneganti, 2015; Karki and Kanneganti, 2019). Therefore, based on the published research, a total of 33 pyroptosis-related signatures were chosen, and 24 genes were retained for subsequent analysis after filtering out the signatures with low expression (sum FPKM value of all samples less than one).

Consensus Cluster Analysis for Pyroptosis-Related Signatures in TNBC

Based on the expression of pyroptosis-related signatures, we performed hierarchical clustering analysis and applied the "ConsensusClusterPlus" R package (Wilkerson and Hayes, 2010) to conduct unsupervised clustering based on Euclidean distance and Ward's linkage methods 1000 repeated times to ensure the classification stability. During the process, the clusters from 2 to 9 were performed, respectively, and the optimal clustering model was determined based on the consensus cumulative distribution function (CDF) plot. Moreover, we performed multiple comparisons among different pyroptosis-subtypes, including for the tumor microenvironment (TME), prognosis, and vital clinical-pathological phenotypes to explore their characteristics. The R packages "survival" (Therneau and Lumley, 2015) and "survminer" (Kassambara et al., 2017) were used to perform Kaplan–Meier survival analysis and draw survival curves between pyroptosis subtypes.

Identification of Differentially Expressed Genes (DEGs) and Functional Enrichment Analysis

To identify the DEGs between pyroptosis subtypes, the empirical Bayesian algorithm was applied through the "Limma" R package (Smyth, 2005), and the significance cutoff was set as adjusted $p < .05$ and absolute fold-change >1. To clarify the biological function and characteristics of pyroptosis clusters, Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was performed by using the "ClusterProfiler" R package (Yu et al., 2012), and the results were visualized using the "ClueGO" plugin in Cytoscape v3.7.1 (Bindea et al., 2009).

TME Cell Infiltration and Gene Set Variation Analysis

To evaluate the immune cell infiltration (ICI) characteristics of TNBC tissues, we used the "CIBERSORT" R package (Chen et al., 2018) to quantitatively analyze the infiltration levels of different immune cells with the LM22 signatures by 1000 random permutations. The tumor purity scores, ICI levels, and stromal contents in different samples were evaluated via the "ESTIMATE" algorithm (Yoshihara et al., 2013). Moreover, through the "c2. cp.kegg.v6.2. symbols" data sets downloaded from the MSigDB database, we performed GSVA using the "GSVA" R package and drew a heatmap to exhibit the different immunogenic pathways (Hanzelmann et al., 2013).

Definition of Immune Characteristics Between High and Low Ps-Score Groups

To further identify a novel index representing the characteristics of the pyroptosis subtypes, we conducted univariate Cox proportional hazards regression analysis for overall survival (OS) to preliminarily screen significant genes through using the "coxph" function in the "survival" R package. Subsequently, to remove the multicollinearity among these candidate genes, LASSO regression was applied to screen independent prognosis-related genes with the optimal penalty parameter and a minimum 10-fold cross-validation (Ranstam and Cook, 2018). After further adjustment, the multivariate Cox regression (stepwise model) was conducted to identify hub genes, and the coefficients obtained from the regression algorithm were used to acquire the Ps-score based on the following formula: $\mathrm{Ps\_score} = \mathrm{val}\,(\mathrm{Gene1}) \times \beta 1 + \mathrm{val}\,(\mathrm{Gene2}) \times \beta 2 + \ldots + \mathrm{val}\,(\mathrm{Gene}\,n) \times \beta \mathrm{n}$. The val (Gene) represents the expression FPKM value of each gene and β the corresponding regression coefficient. Moreover, according to the above formula, the Ps-scores of patients with TNBC were separately calculated, and the patients were divided into high and low subgroups according to the median value as the cutoff value (Sullivan et al., 2004). We also made similar comparisons between high and low Ps-score groups, including the TME, ICI, clinical phenotypes, pyroptosis-related signatures, and correlation of GSVA pathways based on the other four GEO data sets described earlier.

Construction and Evaluation of the Pyroptosis-Related Prediction Model

The multivariate Cox regression (stepwise model) was applied to construct the prognostic model for TNBC-combined Ps-scores and other clinical features, including age, clustering subtypes, clinical stages, and TNM stages. Variables with p-values < .05 were included in the Cox regression model, and the nomogram was further constructed to predict the probability of 1-, 3-, and 5-year survival in patients with TNBC using the "survival" R package. To evaluate and validate the prediction capability of the nomogram, we calculated the concordance index and plotted the calibration curves for 3- and 5-year survival through a bootstrapping method with 1000 resamples. To further investigate the expression of the Ps-score-related genes at the protein level, the Human Protein Atlas (HPA) (Ponten et al., 2008) was used to display the results of the immunohistochemistry (IHC) technique. The detailed information of patients is included in **Supplementary Table S17**.

Exploration of the Significance of Ps-Scores in Response to Clinical Immunotherapy

Based on the IMvigor210 data sets with atezolizumab treatment, we performed a comprehensive comparison between different Ps-score subgroups, including response to immunotherapy, immune phenotype, and clinical remission rate. Moreover, to evaluate the potential therapeutic value of Ps-scores in chemotherapy for TNBC, we calculated the half-maximal inhibitory concentration (IC_{50}) of common chemotherapeutic drugs based on the Genomics of Drug Sensitivity in Cancer (GDSC) databases (Yang et al., 2013). Antitumor drugs such as 5-fluorouracil, cisplatin, docetaxel, doxorubicin, and paclitaxel are recommended for BC treatment by current clinical guidelines. Differences in IC_{50} of these chemotherapeutic drugs between Ps-score subgroups were compared by Wilcoxon test with the results exhibited in box diagrams using the "ggpubr" R package (Whitehead et al., 2019).

Validation of Molecular Subtypes Based on scRNA-Seq Analysis

To validate molecular clusters and further seek biomarkers of each cluster, the Seurat pipeline was selected for subsequent analysis. Using the Seurat package v3.0 (Butler et al., 2018), we transformed the data matrix into a "Seurat object" through the "CreateSeuratObject" R function and performed the necessary quality control. The violin diagram exhibited the number of sequencing reads per sample and the expression of mitochondrial genes (**Supplementary Figure S2A**). Further, to remove the influence from mitochondrial and extreme genes, we kept the number of sequenced genes at 200–10,000, directly including the majority genes, and removed the cells with average gene expression <10 and mitochondrial genes >5%. Then, we conducted data standardization through the "NormalizeData" function with the method of "LogNormalize" and used the top 1500 variable counts to perform PCA using the "RunPCA" function (**Supplementary Figures S2B,C**). Subsequently, t-distributed statistical neighbor embedding (tSNE) was applied to visualize the density clustering, and the "SingleR" package was applied for cell-subtype annotation based on the marker genes (Aran et al., 2019). Moreover, the pseudotime trajectory analysis was further performed using the "Monocle" package v2.0 to expound the potential inner relationship among these cell clusters (Trapnell et al., 2014).

Prediction of Potential miRNA Targets for Prognosis-Associated Signatures

The miRNAs targeting Ps-score-related genes were predicted based on the following databases: TargetScan (http://www.targetscan.org/), starBase (http://starbase.sysu.edu.cn/starbase2/index.php), miRTar (https://mirtarbase.cuhk.edu.cn/), and miRDB (http://www.mirdb.org/). The expression of TNBC miRNA was downloaded from the TCGA-BRCA data sets, and the DEmiRNAs were further identified by the "Limma" R package. Subsequently, we identified the intersection of predicted miRNA by four databases and DEmiRNAs as regulated miRNAs for each hub gene and further visualized the miRNA–mRNA interaction network using Cytoscape v3.7.1.

RESULTS

Overview of Genetic and Biological Characteristics of Pyroptosis-Related Signatures in TNBC

After a series of rigorous screening and quality control steps, a total of 24 pyroptosis-related signatures remained for subsequent analysis in our study (**Supplementary Table S2**). Combined with the normal tissues in GTEx data sets, we first compared the expression of pyroptosis-related signatures between patients with TNBC and normal controls. We found that most pyroptosis-related genes were significantly upregulated in TNBC groups, including *CASP1/3/5/8*, *GSDMA/C*, *NLRC4/P3/P7*, *IL18*, *IL1β*, and *TNF* (**Figure 1A**). PCA indicated that the expression of these pyroptosis-related signatures could be used to divide the TNBC samples and controls into two distinct clusters (**Figure 1B**). In addition, the PCA of internal reference genes revealed that nonsignature genes failed to discriminate TNBC and control cohorts, indicating that the separation created a true distinction based on the pyroptosis-related DEGs in patients with TNBC (**Supplementary Figure S1A**). Moreover, the KEGG functional enrichment analysis revealed that these pyroptosis-related genes were predominantly focused on infectious diseases, immune response, and cellular signal conditioning mechanisms, including the NLR signaling pathway, p53 signaling pathway, TNF signaling pathway, and apoptosis (**Figure 1C**, **Supplementary Table S3**). In terms of genetics, 148 of the 203 samples (72.91%) manifested pyroptosis-related signatures in mutations and the NLR families, especially NLRP3 and NLRP7, exhibiting the highest frequency of mutations (**Figure 1D**, **Supplementary Table S4**). Moreover, the top 10 pyroptosis-related genes with the most frequent mutations were located on the 24 human chromosomes (**Figure 1E**).

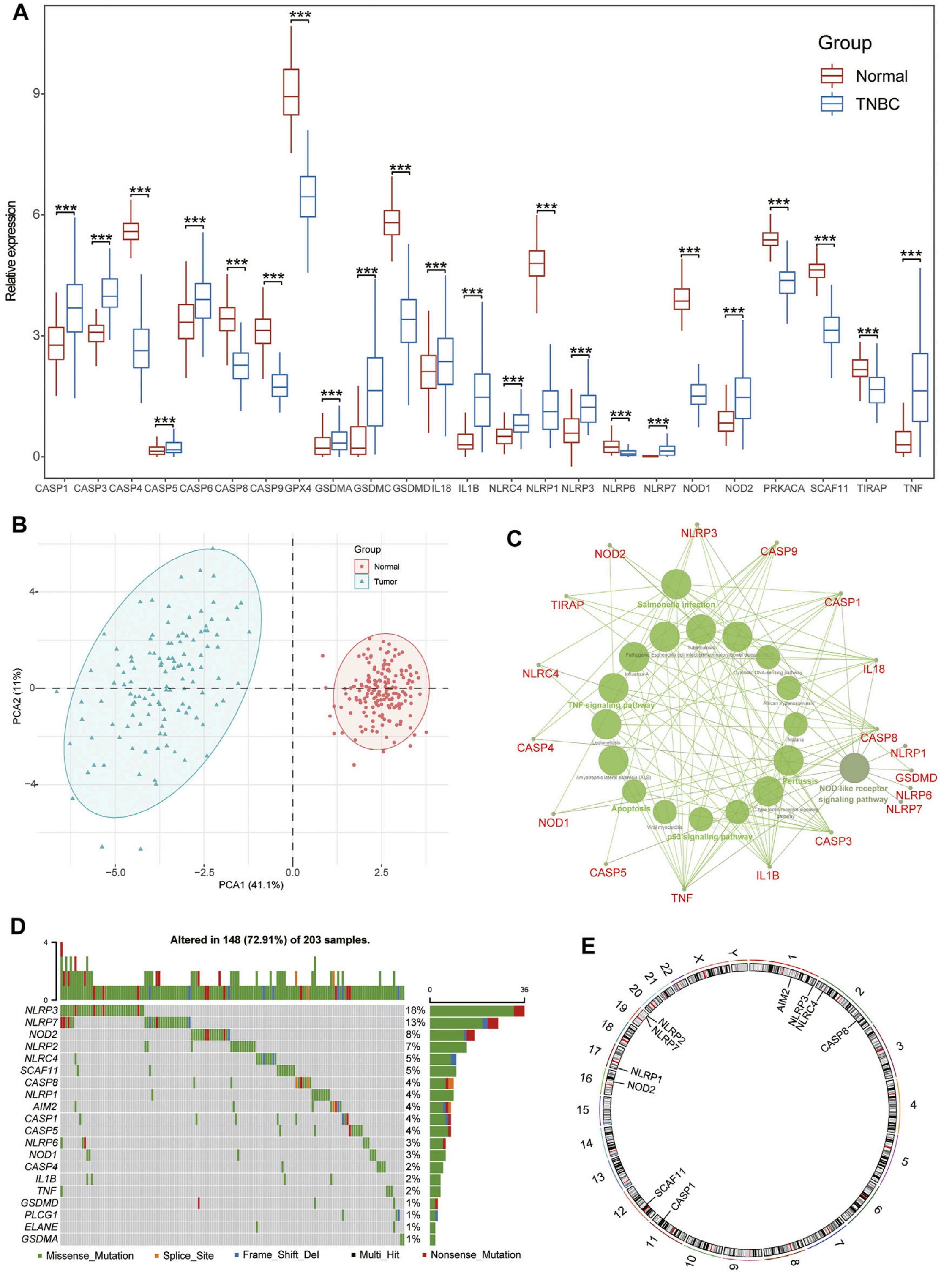


FIGURE 1 | Characteristics of pyroptosis-related signatures in patients with TNBC. **(A)** The expression of pyroptosis-related signatures between normal tissues and TNBC tissues from GTEx and TCGA data sets; Wilcox test, **p* < .05, ***p* < .01, ****p* < .001; ns, not statistically significant. **(B)** PCA showing pyroptosis-related signatures sorted TNBC and control tissues into two clusters. **(C)** The KEGG functional analysis of pyroptosis-related signatures. **(D)** The landscape of mutation profiles in patients with breast cancer from TCGA-BRCA cohort. **(E)** The location of the top 10 pyroptosis-related signatures with the most frequent mutations on the 24 human chromosomes.

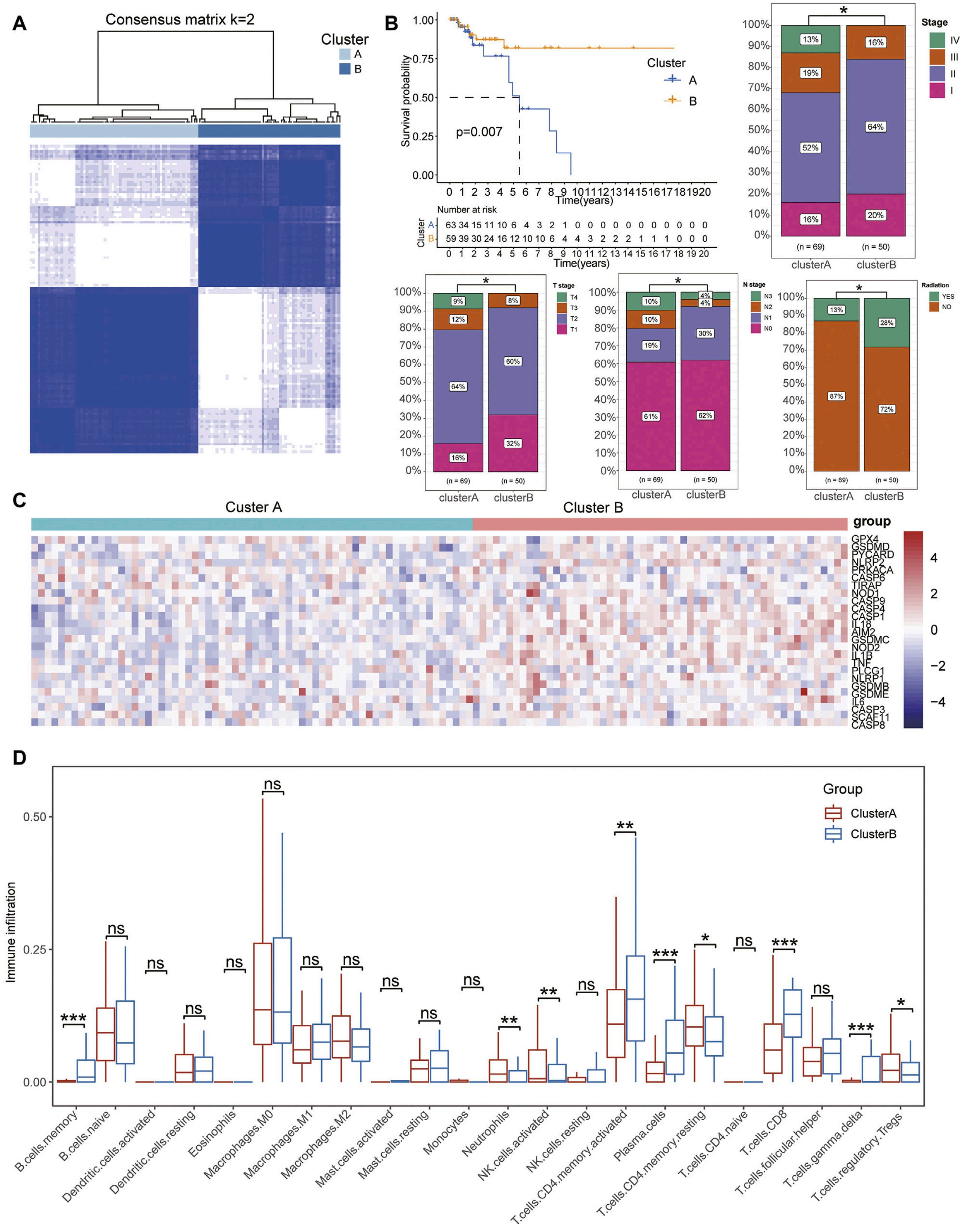


FIGURE 2 | Identification of pyroptosis-related subtypes in TNBC. **(A)** Consensus clustering matrix for *k* = 2 in patients with TNBC. **(B)** Kaplan–Meier curves of OS for patients with

(Continued)

FIGURE 2 | TNBC divided into two subtypes. ClusterB was negatively associated with severe clinical stages, including pathological and TMN stages. **(C)** Heatmap showing the expression of pyroptosis-related signatures upregulated in ClusterB subtypes. **(D)** Boxplots show the difference in immune cell infiltration between ClusterA and ClusterB. Wilcox test, *$p < .05$, **$p < .01$, ***$p < .001$.

Identification of a TNBC Cluster Pattern Based on Pyroptosis-Related Signatures

Based on the expression of pyroptosis-related signatures, we used an unsupervised clustering method to identify the subtypes of patients with TNBC and identified $k = 2$ as the optimum clustering model from $k = 2$ to $k = 9$ clustering with the least area under the consensus CDF curve for 69 patients in clusterA and 50 patients in clusterB (**Figure 2A**, **Supplementary Table S4**; **Supplementary Figure S1B**). To further clarify the intrapatient heterogeneity of patients with TNBC, we performed a comparison in the clinical differences between subtypes and found that patients in clusterB were negatively associated with severe clinical stages, including the pathological and TMN stages. Furthermore, the survival analysis showed that patients in clusterB had a longer median survival time than those in clusterA with more patients in clusterB also receiving radioactive treatments, indicating that patients with TNBC in clusterB might have a better prognosis (**Figure 2B**). There was no significant difference in other clinical indexes including age, sex, M stage, and the ratio of pharmaceutical and surgical therapies (**Supplementary Figure S1C**). Notably, the expression of pyroptosis-related gene signatures was significantly increased in patients in clusterB compared with that of the clusterA cohort (**Figure 2C**). In terms of the immune infiltration scores, adaptive immune response-related lymphocytes (including memory B cells, activated memory CD4$^+$ T cells, plasma cells, CD8$^+$ T cells, and gamma delta T cells) were significantly increased in patients in clusterB compared with the clusterA cohort. However, innate immunity and immunoregulation-related cells were significantly infiltrated in clusterA cohorts, including neutrophils, activated natural killer cells, resting memory CD4$^+$ T cells, and regulatory T cells (Tregs) (**Figure 2D**, **Supplementary Table S5**). Higher stromal scores and immune scores with lower tumor purity were also detected in patients in clusterB compared with the clusterA groups (**Figures 3A,B**).

Identification of DEGs Based on Pyroptosis-Related Clusters

Considering the biological characteristics of immune subtypes in TNBC, we conducted a DEG analysis between the two subtypes. Through comparing clusterA with clusterB groups, a total of 262 DEGs (including 13 clusterA- and 249 clusterB-related genes) in TNBC were identified (**Figure 3C**, **Supplementary Table S6**). To further interpret biological processes and pathways of pyroptosis-related subtypes, these DEGs were chosen to perform KEGG functional analysis. The results showed that clusterA-related genes were not enriched in any significant pathways while the clusterB-related signatures were predominantly enriched in immune activation-associated pathways, including natural killer cell–mediated cytotoxicity, the toll-like receptor signaling pathway, chemokine signaling pathway, cytokine–cytokine receptor interaction, NF-kappa B signaling pathway, TNF signaling pathway, Th17 cell differentiation, and Th1 and Th2 cell differentiation (**Figure 3D**, **Supplementary Table S7**).

Development of Ps-Score and Characteristic of Ps-score-related Subgroups

After successively including the 262 DEGs in univariate Cox regression, LASSO regression, and multivariate Cox regression analysis as candidate prognosis-associated genes, we identified six hub genes (including *CFB*, *IFITM1*, *EPSTI1*, *MARCO*, *CXCL13*, and *CCL5*) from the Ps-score signatures based on their β coefficients (**Figures 3E,F**, **Supplementary Table S8**). In addition, the expression of these hub genes was higher in clusterB subgroups. Based on the IHC data from the HPA database, the expression of these hub genes at the protein level was further validated in BC, especially CFB, ESPIT1, and IFITM1 (**Figure 3G**, **Figure 6C**). Based on the expression of these genes and their corresponding β coefficients, the Ps-score was defined by the following formula: $\text{Ps\_score} = -0.365 \times CFB - 0.45 \times IFITM1 - 0.298 \times EPSTI1 - 0.461 \times MARCO - 0.26 \times CXCL13 + 0.439 \times CCL5$ (**Supplementary Table S9**, the gene name represents the corresponding gene expression FPKM values). Subsequently, those patients with TNBC were divided into a high and low Ps-score subgroups with median value (−5.28) as the cutoff; the high Ps-score cohorts exhibited a worse prognosis than that of low Ps-score patients in the TCGA data sets (**Figure 4A**). To prove the universal value of the Ps-score in TNBC, we also performed survival analysis of this score in four extrinsic GEO cohorts and obtained the same results (**Figure 4B**).

To investigate the biological characteristics of the Ps-scores, we compared the expression of pyroptosis-related genes and ICI between different Ps-score groups and further explored the correlation between significant clinical phenotypes and the Ps-scores. The results reveal that clusterB possessed a lower level of Ps-scores associated with pharmaceutical and surgical therapy as well as lower pathological stages in patients with TNBC (**Figure 4**C&E). Interestingly, the expression of pyroptosis-related signatures was significantly increased in the low Ps-score groups, including the CASP, GSDM, and NOD families, as well as inflammatory factors, suggesting the potential role of pyroptosis activation in the low-score of TNBC cohorts with better prognosis (**Figure 4D**). Moreover, ROC analysis showed 1-, 3-, and 5-year AUC values of the Ps-scores for predicting the prognosis of patients with TNBC of 0.867, 0.867, and 0.906, respectively, in the TCGA sets (**Figure 5A**). Furthermore, immune infiltration analysis revealed that substantial immune cells were significantly inhibited in the high Ps-score groups, including that of CD8$^+$ T cells, follicular helper T cells, activated CD4$^+$ memory T cells, and plasma cells (**Figure 5D**). The correlation analysis also indicates that the Ps-scores were significantly positively associated with the levels of tumor purity ($R = 0.57$, $p < .001$) but negatively

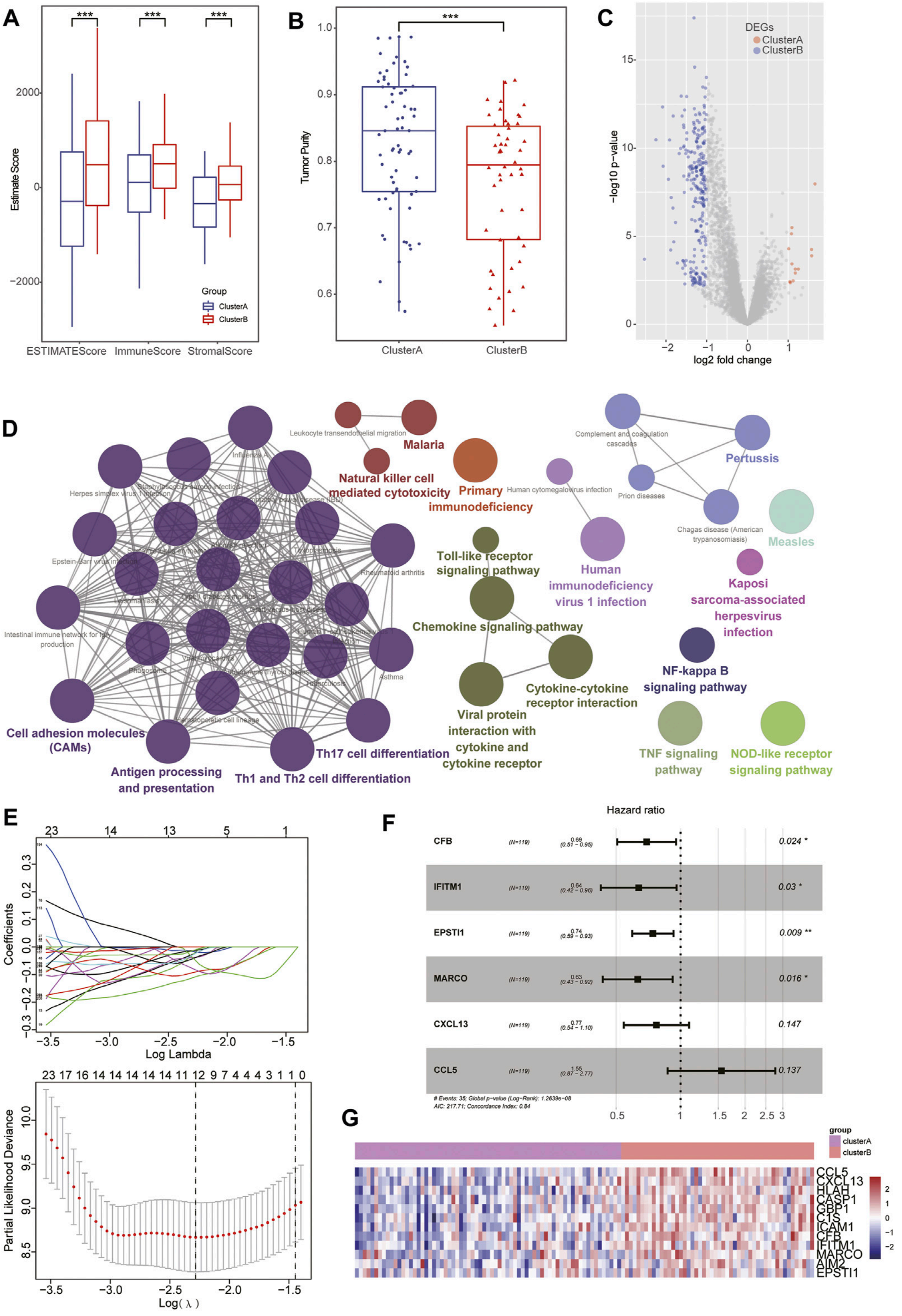


FIGURE 3 | Identification of DEGs based on pyroptosis-related clusters. **(A–B)** Box plot showing higher stromal and immune scores with lower tumor purity detected in patients of clusterB than the clusterA group. **(C)** Volcano plots displaying the up- and downregulated DEGs between two subgroups in TNBC cohorts. **(D)** Bubble diagram showing the results of KEGG enrichment analysis of the subtypes. **(E)** LASSO coefficient profiles of 12 prognostic related genes and 10-times cross-validation for tuning parameter selection in the LASSO model. **(F)** Forest map displaying the HR and *p*-value of six hub genes after multivariate Cox regression analysis. **(G)** Heatmap showing the distinct expression of six hub genes between pyroptosis-related clusters.

associated with stromal and immune scores ($R = -0.48$, $p < .001$ and $R = -0.50$, $p < .001$, respectively) (**Figure 5E**).

Based on the Ps-scores and some primary clinical characteristics, multivariate Cox regression analysis was conducted to construct a nomogram that could accurately predict the probability of the 1-, 3-, and 5-year survival for patients with TNBC. The Ps-scores, age, pyroptosis-related cluster, and clinical stages were considered as related predictors for the prognosis of patients with TNBC and incorporated into the nomogram with significant regression coefficients and *p*-values (**Figure 5B**, **Supplementary Table S10**). From the nomogram, we could observe that the Ps-score contributed the most to the total score with a 0.716 concordance index (**Figure 5B**, **Supplementary Table S11**). Calibration curves exhibited that the nomogram had a good prediction capacity in both 3- and 5-year OS for patients with TNBC (**Figure 5C**).

Significance of Ps-Scores in the Prediction of Response to Immunotherapy and Common Chemotherapeutics

The alluvial diagram visualized the status changes in the different characteristics of patients (**Figure 6A**). We found that patients with a low Ps-score in clusterB subtypes had a higher ratio of survival status. When using the TCGA and other external GEO data sets, the results of GSVA demonstrated the coincident negative correlation between Ps-scores and immunoregulation-related pathways, such as the toll-like receptor signaling pathway, antigen processing and presentation, rig I-like receptor signaling pathway, T cell receptor signaling pathway, NOD-like receptor signaling pathway, and JAK-STAT signaling pathway (**Figure 6B**, **Supplementary Table S12**).

To further explore the role of Ps-scores in predicting the therapeutic benefit in TNBC, we first calculated the Ps-scores of patients who accepted anti-PD-L1 immunotherapy from the GSE157284 and IMvigor210 cohorts before assigning them into high or low Ps-score groups. From the GSE157284 data set, patients with an effective response rate to anti-PD-L1 therapy showed lower Ps-scores, whereas the relative expression of immune-checkpoint genes was significantly increased in the high Ps-scores groups (**Figures 6D,E**, **Supplementary Table S13**). Congruously, we also found that the low Ps-score group had higher expression of immune-checkpoint genes in the IMvigor210 cohort and effective anti-PD-L1 responders also exhibited lower Ps-scores (**Figures 6F,G**). In metastatic urothelial cancers of the IMvigor210 data sets, distinct immunological subtypes might result in opposite therapeutic responses. Therefore, we also compared the Ps-score levels among these subtypes and found that the lowest Ps-score was in the inflamed subtypes (**Figure 6H**). Moreover, the rate of complete remission (CR) after immunotherapy was also increased in the low Ps-score cohort compared with the high Ps-score cohort, with the low Ps-score group validated to have a better prognosis for metastatic urothelial cancers (**Figures 6I,J**). All these results suggested that the Ps-score might serve as a significant indicator in immunotherapy decision making for cancers.

Besides checkpoint blocker therapy, we also investigated the potential associations between the Ps-scores and the curative efficacy of common chemotherapeutics in treating BCs. The IC_{50} value was calculated for five common anti-BC chemotherapeutic drugs obtained from the GDSC databases, including 5-fluorouracil, cisplatin, docetaxel, doxorubicin, and paclitaxel (**Supplementary Table S14**). Notably, all the drugs exhibited lower IC_{50} values in the low Ps-score groups, indicating patients with low Ps-scores might obtain a better curative efficacy from common chemotherapy (**Figure 6K**). Collectively, these outcomes indicate that Ps-scores could be associated with the response to immunotherapy and common chemotherapy.

Validation of Cellular Subtypes in TNBC Through scRNA-Seq Analysis

To validate the potential subtypes of patients with TNBC, the GSE118389 data set along with 1534 cells from six TNBC tissues were used to identify concrete cellular subtypes and corresponding marker genes. A total of 12 distinct cellular clusters were identified through tSNE analysis (**Figure 7A**) with the marker-genes of each cluster listed in **Supplementary Table S15**. Moreover, the results of cell-type annotation using the "SingleR" package indicated these cell clusters fell into six cellular subtypes, including epithelial cells, erythrocytes, $CD8^+$ T cells, fibroblasts, endothelial cells, and monocytes, of which the epithelial cells were the most common cell type with six subtypes (**Figure 7B**). Notably, of the six Ps-score-related genes, five genes were identified as marker genes, and their expression in each cellular subtype is shown in a bubble diagram (**Figure 7C**). The expression of *CCL5* and *IFITM1* were generally increased in nearly all cellular subtypes and significantly higher than the expression of other signatures. Although the expression of remanent genes was relatively lower in these cells, significant cellular specificity was found in these Ps-score-related genes. For example, *MARCO* was particularly expressed in the epithelial cell subtype 4 and monocytes while *CFB* was particularly expressed in the fibroblast subtype 2 as well as epithelial cell subtypes 3 and 4. Interestingly, *EPSTI1* was significantly expressed in immune-related cellular subtypes including monocytes and $CD4^+$ T cells, consistent with the results of IHC. In addition, the pseudotime trajectory analysis also revealed the arrangement of different cellular subtypes that formed a certain rule based on its spatial relationships (**Supplementary Figure S2E**). Concretely, the trajectory analysis of epithelial cells revealed that a small quantity of epithelial cell subtype 3 was distributed at the start of the trajectory while mixed cells from epithelial cell subtypes 5 and 6 were distributed at the end of the trajectory. Moreover, epithelial cell subtype 2 was uniformly located behind epithelial cell subtypes 1, whereas epithelial cell subtype 4 nearly existed throughout the trajectory (**Figure 7D**).

Prediction and Validation of miRNAs Interacted With Hub Ps-Score Genes

To further explore the potential regulatory role of miRNAs targeted to these Ps-score-related genes in TNBC, we found 223 probable miRNAs and successfully constructed the

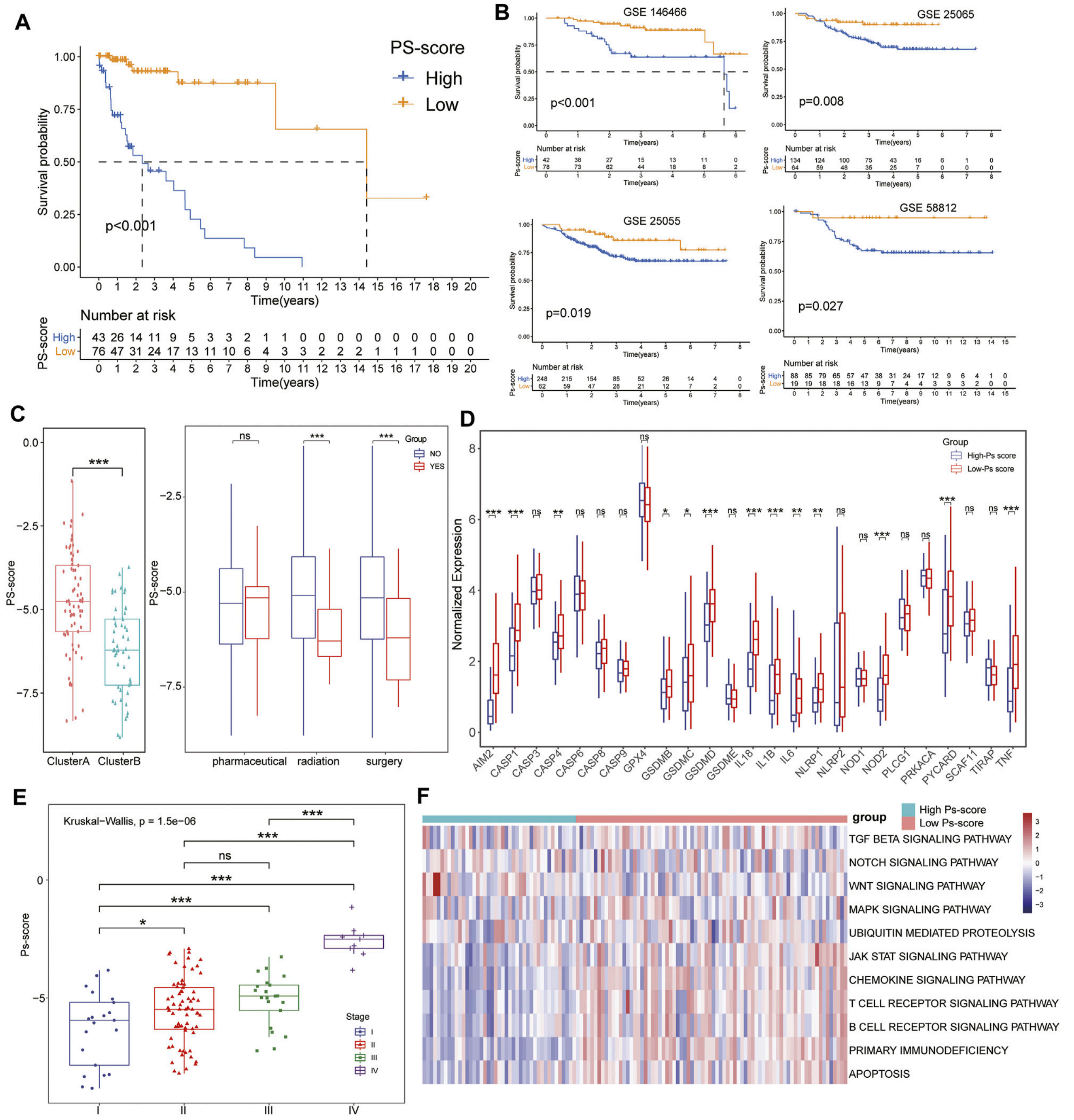


FIGURE 4 | Characteristics of Ps-scores and correlation to pyroptosis in TNBC. **(A–B)** Kaplan–Meier curves of overall survival (OS) for the patients with TNBC in the high- and low-Ps-score groups in the TCGA **(A)** or GEO data sets **(B)**. **(C, E)** The box plots show that ClusterB possessed a lower level of Ps-scores associated with pharmaceutical and surgical therapies, and lower pathological stages. **(D)** Box diagram displaying the expression of pyroptosis-related signatures significantly increased in the low Ps-score groups. **(F)** Heatmap used to visualize pathways analyzed by GSVA showing the active biological pathways in distinct pyroptosis-related clusters.

miRNA-hub gene interaction network based on the prediction of the TargetScan, starBase, miRTar, and miRDB databases (**Figure 7E**, **Supplementary Table S16**). In addition, to validate the regulatory role of miRNAs in TNBC, we also identified 37 DEmiRNAs between pyroptosis-related clusters, before screening the top 20 miRNAs serving as vital regulatory factors (**Figures 7F,G**). Finally, we successfully simplified the interaction network with 20 miRNAs and four targeted genes to verify that *ESPIT1* and *CXCL13* are the most active targets regulated by massive DEmiRNAs (**Figure 7H**).

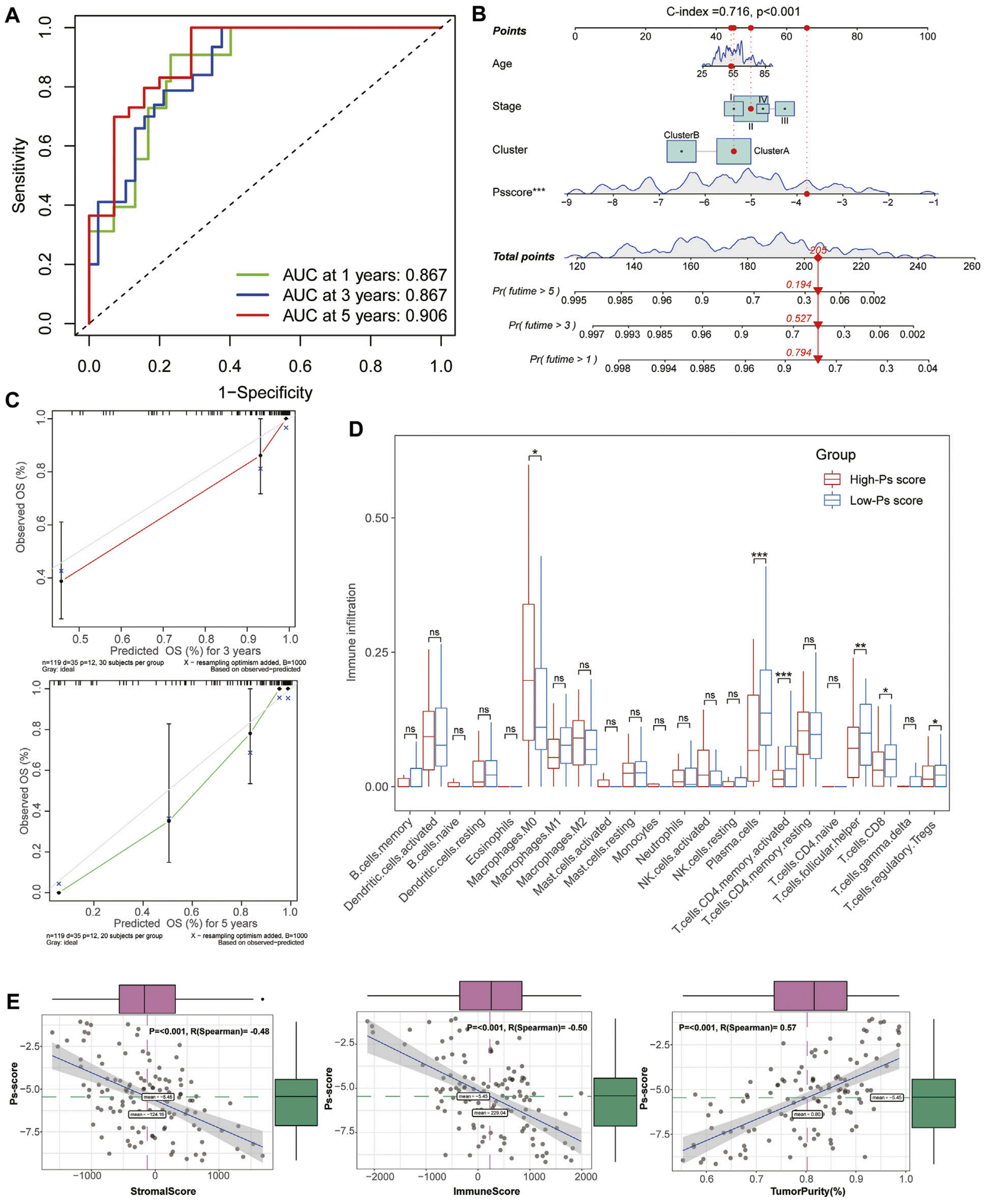


FIGURE 5 | Establishment and evaluation of the Ps-score-related prognostic model for patients with TNBC. **(A)** Time-dependent receiver operating curves of 1-, 3-, and 5-year survival for patients with TNBC using Ps-scores. **(B)** Combined nomogram for predicting the probability of 1-, 3-, and 5-year survival for patients with TNBC. **(C)** Calibration curve of the established nomogram with 3- and 5-year survival, respectively. **(D)** Immune infiltration analysis revealed that substantial immune cells were significantly inhibited in high Ps-score groups. **(E)** Correlation analysis shows the Ps-scores significantly positively associated with the levels of tumor purity (*R* = 0.57) and negatively associated with stromal and immune scores (*R* = −0.48 and *R* = −0.50, respectively).

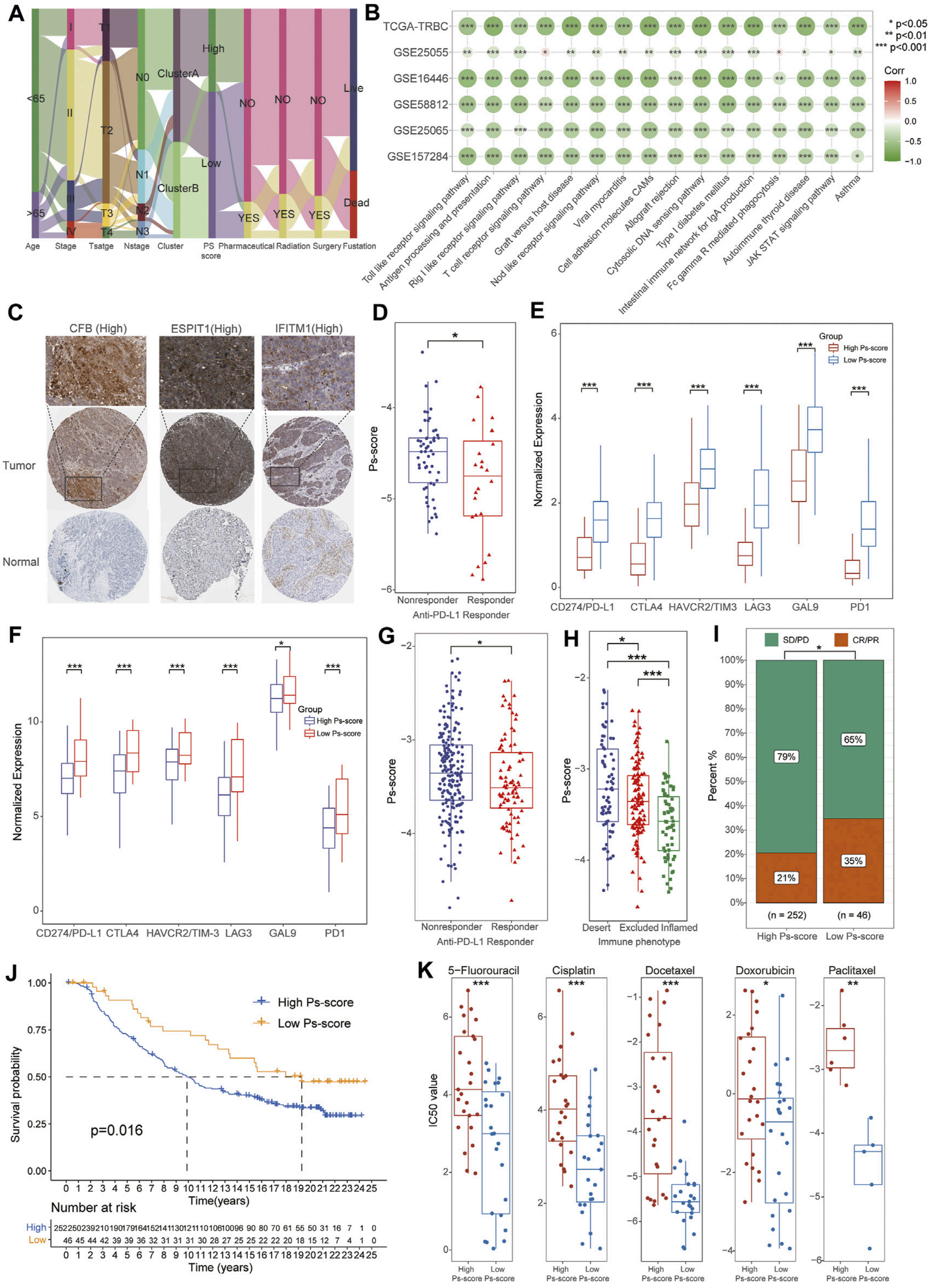


FIGURE 6 | The significance of Ps-scores in the prediction of response to immunotherapy and common chemotherapeutics for TNBC. **(A)** Alluvial diagram visualizing the status changes from different characteristics of patients. **(B)** Correlation analysis of Ps-scores and GSVA pathways in TCGA and GEO data sets. The size of nodes represents the correlation coefficient, and the red or green represents positive or negative correlation, respectively. **(C)** The immunohistochemical results of Ps-score-related genes from the HPA database. **(D)** The box plots displayed patients with better responses to PD-L1 treatment exhibited higher Ps-scores using GSE157284. **(E)** The relative expression of immune-checkpoint genes was significantly increased in the high Ps-scores groups of GSE157284 data sets. **(F)** The relative expression of immune-checkpoint genes was significantly increased in the high Ps-scores groups of IMvigor210 cohorts. **(G)** The box plots indicated patients with better

(Continued)

FIGURE 6 | responses to PD-L1 treatment exhibited higher Ps-scores using IMvigor210 cohorts. **(H–I)** Ps-score in different ACRG subtypes and the rate of CR after immunotherapy in IMvigor210 cohorts. **(J)** Kaplan-Meier curves of OS for the patients with TNBC in high and low Ps-score groups of IMvigor210 data sets. **(K)** Comparison of IC_{50} value of 5-fluorouracil, cisplatin, docetaxel, doxorubicin, and paclitaxel in high and low Ps-score groups using GDSC databases.

DISCUSSION

As a malignant tumor with high mortality, TNBC is known for its poor prognosis, which stems from ineffective therapeutic response to immunotherapy due to tumor biological heterogeneity. Recently, the IMpassion130 trial demonstrated that the combination of atezolizumab, a PD-L1 inhibitor, and nab-paclitaxel could prolong the OS in PD-L1- patients with positive advanced TNBC, heralding the emergence of immunotherapy as an effective treatment strategy for TNBC (Schmid et al., 2018). In addition, The U.S. Food and Drug Administration (FDA) also approved SP142, a PD-L1 IHC assay, as an auxiliary test to identify eligibility for atezolizumab therapy in patients with advanced TNBC. However, the results of IHC staining are still short of high interlaboratory reproducibility with subjective judgment. For example, the IHC levels of PD-L1 were investigated in a total of 443 patients with BC by (Wang et al., 2017) but only ~16% of these tumors exhibited positive PD-L1 levels. Although the PD-L1 test plays a potentially significant role in the management of multiple advanced carcinomas, objective standardization for this test has not been achieved; hence, its current use in the clinical practice poses a twofold risk to patients: false positive could result in potentially toxic therapies resulting in unforeseen complications, such as miscarriages, or PD-L1 false negatives would benefit from therapy but are excluded from receiving treatment (Reisenbichler et al., 2020). Therefore, identification of novel immunotherapy-related subtyping and reliable objective prognostic indicators for immunotherapy in TNBC is urgently needed.

In contrast to apoptosis, pyroptosis usually occurs in abnormal cells infected by microbes as a positive programmed cell death process, thus inducing the release of pro-inflammatory cytokines and activating an inflammatory response (Bedoui et al., 2020). Prompted by microbes, pyroptosis can also be converted from apoptosis and play various roles in multiple tumors. Pyroptosis has shown antitumor effects through inhibiting the tumor growth in liver and gastric cancers while showing both suppression or promotion effects in BC (Zaki et al., 2010; Chen et al., 2012; Shao et al., 2021). Shi et al. (2015) demonstrated that the activation of the NLRP3 (NOD-, LRR-, and pyrin domain-containing 3) inflammasome was integral for the activation of pyroptosis by recruiting CASP1, further leading to cleavage of GSDMD. In the present study, we explored all the signatures and pathways directly related to pyroptosis in TNBC and detected that the NLRP3/CASP1/GSDMD pathway-related pyroptosis was activated in patients with TNBC, implying that pyroptosis might participate in the mechanism of TNBC, which was associated with the prognosis of TNBC.

The classification of patients based on pathognomonic gene expression profiles is considered a proven method and applied to various studies of TNBC, including autophagy-related signatures (Kim et al., 2012), N6-methyladenosine (Wu et al., 2021), and immune cell infiltration (Harano et al., 2018). In this study, we first proposed a pyroptosis-related molecular subtype based on clustering pyroptosis-related signatures with distinct clinical and immunological characteristics. Interestingly, the characteristics of the two molecular subtypes manifested in significant homogeneity. We detected that patients in ClusterB presented a longer median survival time than those in ClusterA, whereas patients in ClusterB also negatively associated with serious clinical stages, including pathological and TMN stages, suggesting these pyroptosis-related signatures were also significantly associated with different survival risks in patients with TNBC. Our results also reached some consensus: (1) nearly all the pyroptosis-related signatures exhibited higher expression in patients in ClusterB; (2) ClusterB was a specific subtype with a better prognosis and slighter clinical pathological phenotypes; (3) ClusterB was identified as an immune-activated phenotype with higher TME immune scores and infiltration levels of adaptive immune response-related immune cells.

To further explore the potential biological functional features of the pyroptosis-related subtypes in TNBC, we investigated the DEGs between the two subtypes and performed KEGG function enrichment analysis. Consistent with the immunological signatures of subtypes, functional enrichment analysis revealed that immune-activation associated pathways, including natural killer cell-mediated cytotoxicity, the toll-like receptor signaling pathway, chemokine signaling pathway, cytokine-cytokine receptor interaction, NF-kappa B signaling pathway, TNF signaling pathway, and Th17 cell differentiation as well as Th1 and Th2 cell differentiation were significantly enriched in the ClusterB cohorts. Of these pathways, the activation of multiple immune pathways is reported to suppress metastatic spread in TNBC (Zanker et al., 2020) and could be the potential mechanism for a better prognosis of patients in Cluster B.

Furthermore, to increase the clinical application value and create better clinical practicability, we successfully constructed a novel pyroptosis-related scoring tool (Ps-score) to determine the prognostic risk of TNBC based on six hub genes from two clusters. High-expression of these risk signatures at the protein levels was confirmed by IHC from the HPA database, and the Ps-scores effectively stratified patients with TNBC from the TCGA and GEO data set, respectively, into high- and low-risk groups. Survival analysis revealed that the low-score groups had longer OS than patients with high Ps-scores, and ROC curves exhibited a great predictive capacity of Ps-scores for the 1-, 3-, and 5-year survival of TNBC. In addition, the Ps-scores were significantly decreased in the ClusterB cohorts, indicating that the Ps-score could reflect the heterogeneity of patients with TNBC. Moreover, the Ps-score also represented patients with different clinical outcomes and was associated with the response to

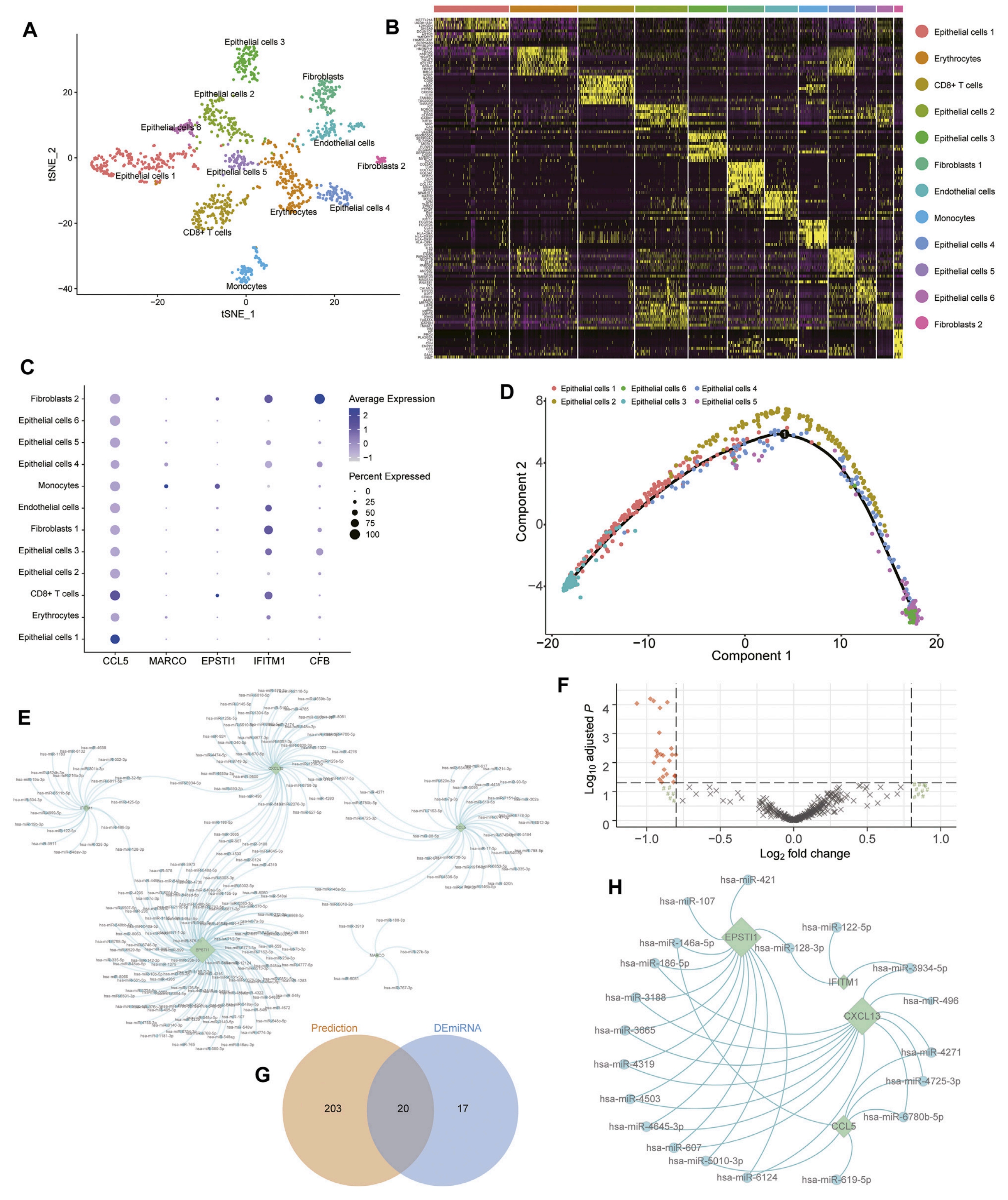


FIGURE 7 | Validation of cellular subtypes in TNBC through scRNA-seq analysis. **(A)** Visualization of tSNE colored according to cell types for TNBC single-cell transcriptomes. **(B)** Heatmap revealing the scaled expression of top 10 marker-genes for each cluster defined in **(A)**. **(C)** Dot plots showing the expression of indicated Ps-score-related genes for each cell cluster. **(D)** Monocle pseudospace trajectory analysis revealing the cellular lineage progression of Epithelial cells subtypes in patients with TNBC colored according to different Epithelial cell clusters. **(E)** The miRNA-hub genes interaction network based on the prediction of four databases. The round node represented the miRNA and the rhomb represented the Ps-score-related genes. **(F)** Volcano plots displayed the upregulated and downregulated DEmiRNAs between two subgroups in patients with TNBC. **(G)** Venn diagram of the predicted miRNA and DEmiRNAs.

immunotherapy. The patients with high Ps-scores exhibited worse clinical prognosis and lower expression of immune checkpoints. TME ICI analysis also demonstrated that the Ps-score was significantly negatively correlated with the infiltration levels and tumor purity, suggesting its value in immunotherapy. Finally, a correlation analysis of the Ps-score and pathways based on GSVA results demonstrate a coincident negative relation in both TCGA and other GEO data sets, implying the activation of the various immune-related processes might occur more frequently in low Ps-score cohorts. Notably, besides immune-related pathways activated in low-score patients, massive activation of signal regulatory pathways was observed in high Ps-score groups, including TGF-β1, Wnt, Notch, and the MAPK signaling pathway, which is reportedly involved in the mechanism of TNBC and as target pathways for drug treatments (Giltnane and Balko, 2014; Kim et al., 2016; Pohl et al., 2017; Giuli et al., 2019).

Furthermore, using the IMvigor210 data sets, we also speculated that the Ps-score might be applicable to estimate the clinical response to immunotherapy in other tumors as well. Besides immunotherapy, common chemotherapeutic drugs also demonstrated lower IC_{50} values in the low Ps-score cohorts, including 5-fluorouracil, cisplatin, docetaxel, doxorubicin, and paclitaxel from the GDSC database, implying that these chemotherapeutic drugs would be more effective in patients with TNBC with low Ps-scores. Overall, these findings from external data sets validated the potential benefits of using the Ps-score system and indicated its role in predicting curative responses to common chemotherapies and immune checkpoint therapies.

Finally, the scRNA-seq analysis demonstrated the authentic existence of cellular subtypes with their marker genes in patients with TNBC and clearly showed the distribution of Ps-score-related genes in each subtype. Admittingly, TNBC originated from epithelial cells, and the results of the scRNA-seq also demonstrated multiple subtypes of epithelial cells, reflecting different clusters of tumor cells. Interferon inducible transmembrane 1 (IFITM1) is reported to promote the progression of TNBC through regulating *integrin*, *NFκB*, and *IL6* gene expression and might serve as a novel therapeutic target for patients with *IFITM1*$^+$ TNBC (Provance et al., 2021). Of the Ps-score-related genes, our analysis also detected that *IFITM1* exhibited relatively high expression in epithelial cell subtypes 3–5, consistent with the above patients with *IFITM1*$^+$ TNBC. In previous studies, the epithelial-stromal interaction 1 (EPSTI1) is also shown to modulate the extrinsic apoptotic pathway in TNBC cell lines, which highlighted its potential as a therapeutic target for patients with TNBC (Capdevila-Busquets et al., 2015). Interestingly, *EPSTI1* is overexpressed in monocytes and CD8$^+$ T cells, suggesting EPSTI1 might participate in the process of extrinsic apoptosis with the activation of the immune response. Moreover, the pseudotime trajectory analysis displayed the distribution of tumor epithelial subtypes and demonstrated the existence of inner heterogeneity and potential cellular differentiation in patients with TNBC. For the common subtypes of BCs, microRNA profiles from different breast cells were applied to distinguish and reflect different subtypes, including luminal A, luminal B, and basal and malignant myoepithelioma, indicating that the expression of genes in cells could directly reflect the different subtypes in BCs (Bockmeyer et al., 2011). Despite the differences in cellular and individual subtypes, pyroptosis-related signature genes distinguished both subtypes of TNBC based on their differential expression. Combined with the differential expression of miRNAs between pyroptosis-related clusters, we ultimately constructed a miRNA–mRNA interaction network, including 20 miRNAs and four hub genes and found that *EPSTI1* and *CXCL13* were the central nodes with the most miRNA regulation.

Our study has the limitation that the high-throughput sequencing data sets for initial analysis were relatively insufficient as it was simply obtained from public databases. The corresponding results and conclusions remain to be further investigated through more external congeneric research and should be validated via functional experiments *in vivo* and *in vitro*. Furthermore, several conclusions of this study require further research to confirm its reproducibility, improve the clinical application of pyroptosis-related clusters, and elaborate on the role of Ps-scores in predicting the response to immunotherapy for TNBC.

CONCLUSION

Our study is the first to propose molecular subtypes based on clustering pyroptosis-related signature expression with distinct clinical and immunological signatures in patients with TNBC. Moreover, we identified and validated a Ps-score system as an effective tool to predict the OS and immunotherapy efficacy in patients with TNBC. Finally, we preliminarily explored the cellular subtypes using scRNA-seq data sets to demonstrate the heterogeneity of TNBC and successfully construct an interaction network to expound the regulatory miRNA targeted Ps-score-related signatures. The various transcriptomic analyses facilitated the screening of significant genetic signatures of TNBC to provide a new clinical application of Ps-scores in predicting prognosis and chemo-immunotherapeutic response for patients with TNBC.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

CL and JP contributed to data acquisition, analysis and drafting of the manuscript. YJ and YW participated in the process of data acquisition and analysis. ZJ contributed to figures presentation and revision of the manuscript. XC contributed to the design of the study. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the General Scientific Projects of the Zhejiang Education Department (Y202147905).

ACKNOWLEDGMENTS

We would like to acknowledge the TCGA, GEO, GTEx, and HPA database for providing data. We would also like to acknowledge Dr. Jing Luo (Rheumatology Department, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou; Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing, China) for the excellent suggestions in statistics and technical assistance of this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.788670/full#supplementary-material

REFERENCES

Aran, D., Looney, A. P., Liu, L., Wu, E., Fong, V., Hsu, A., et al. (2019). Reference-Based Analysis of Lung Single-Cell Sequencing Reveals a Transitional Profibrotic Macrophage. *Nat. Immunol.* 20, 163–172. doi:10.1038/s41590-018-0276-y

Bedoui, S., Herold, M. J., and Strasser, A. (2020). Emerging Connectivity of Programmed Cell Death Pathways and its Physiological Implications. *Nat. Rev. Mol. Cel. Biol.* 21, 678–695. doi:10.1038/s41580-020-0270-8

Bergin, A. R. T., and Loi, S. (2019). Triple-negative Breast Cancer: Recent Treatment Advances. *F1000Res* 8, F1000 Faculty Rev–1342. doi:10.12688/f1000research.18888.1

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape Plug-In to Decipher Functionally Grouped Gene Ontology and Pathway Annotation Networks. *Bioinformatics* 25, 1091–1093. doi:10.1093/bioinformatics/btp101

Bockmeyer, C. L., Christgen, M., Müller, M., Fischer, S., Ahrens, P., Länger, F., et al. (2011). MicroRNA Profiles of Healthy Basal and Luminal Mammary Epithelial Cells Are Distinct and Reflected in Different Breast Cancer Subtypes. *Breast Cancer Res. Treat.* 130, 735–745. doi:10.1007/s10549-010-1303-3

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating Single-Cell Transcriptomic Data across Different Conditions, Technologies, and Species. *Nat. Biotechnol.* 36, 411–420. doi:10.1038/nbt.4096

Capdevila-Busquets, E., Badiola, N., Arroyo, R., Alcalde, V., Soler-Lopez, M., and Aloy, P. (2015). Breast Cancer Genes PSMC3IP and EPSTI1 Play a Role in Apoptosis Regulation. *PLoS One* 10, e0115352. doi:10.1371/journal.pone.0115352

Carithers, L. J., and Moore, H. M. (2015). The Genotype-Tissue Expression (GTEx) Project. *Biopreserv. Biobank.* 13, 307–308. doi:10.1089/bio.2015.29031.hmm

Chen, L. C., Wang, L. J., Tsang, N. M., Ojcius, D. M., Chen, C. C., Ouyang, C. N., et al. (2012). Tumour Inflammasome-Derived IL-1β Recruits Neutrophils and Improves Local Recurrence-Free Survival in EBV-induced Nasopharyngeal Carcinoma. *EMBO Mol. Med.* 4, 1276–1293. doi:10.1002/emmm.201201569

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12

Giltnane, J. M., and Balko, J. M. (2014). Rationale for Targeting the Ras/MAPK Pathway in Triple-Negative Breast Cancer. *Discov. Med.* 17, 275–283.

Giuli, M. V., Giuliani, E., Screpanti, I., Bellavia, D., and Checquolo, S. (2019). Notch Signaling Activation as a Hallmark for Triple-Negative Breast Cancer Subtype. *J. Oncol.* 2019, 8707053. doi:10.1155/2019/8707053

Global Burden of Disease Cancer; CollaborationFitzmaurice, C., Allen, C., Barber, R. M., Barregard, L., Bhutta, Z. A., et al. (2017). Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived with Disability, and Disability-Adjusted Life-Years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol.* 3, 524–548. doi:10.1001/jamaoncol.2016.5688

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7

Harano, K., Wang, Y., Lim, B., Seitz, R. S., Morris, S. W., Bailey, D. B., et al. (2018). Rates of Immune Cell Infiltration in Patients with Triple-Negative Breast Cancer by Molecular Subtype. *PLoS One* 13, e0204513. doi:10.1371/journal.pone.0204513

He, Y., Jiang, Z., Chen, C., and Wang, X. (2018). Classification of Triple-Negative Breast Cancers Based on Immunogenomic Profiling. *J. Exp. Clin. Cancer Res.* 37, 327. doi:10.1186/s13046-018-1002-1

Hoffman-Censits, J. H., Grivas, P., Van Der Heijden, M. S., Dreicer, R., Loriot, Y., Retz, M., et al. (2016). IMvigor 210, a Phase II Trial of Atezolizumab (MPDL3280A) in Platinum-Treated Locally Advanced or Metastatic Urothelial Carcinoma (mUC). *Am. Soc. Clin. Oncol.* 34, 355. doi:10.1200/jco.2016.34.2_suppl.355

Karki, R., and Kanneganti, T.-D. (2019). Diverging Inflammasome Signals in Tumorigenesis and Potential Targeting. *Nat. Rev. Cancer* 19, 197–214. doi:10.1038/s41568-019-0123-y

Kassambara, A., Kosinski, M., and Biecek, P. (2017). Package 'Survminer' [J]. *Drawing Survival Curves Using 'ggplot2'.* (R Package Version 03 1).

Kim, S., Jung, W. H., and Koo, J. S. (2012). Differences in Autophagy-Related Activity by Molecular Subtype in Triple-Negative Breast Cancer. *Tumor Biol.* 33, 1681–1694. doi:10.1007/s13277-012-0424-1

Kim, S., Lee, J., Jeon, M., Lee, J. E., and Nam, S. J. (2016). Zerumbone Suppresses the Motility and Tumorigenecity of Triple Negative Breast Cancer Cells via the Inhibition of TGF-β1 Signaling Pathway. *Oncotarget* 7, 1544–1558. doi:10.18632/oncotarget.6441

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments. *Bioinformatics* 28, 882–883. doi:10.1093/bioinformatics/bts034

Lin, N. U., Vanderplas, A., Hughes, M. E., Theriault, R. L., Edge, S. B., Wong, Y.-N., et al. (2012). Clinicopathologic Features, Patterns of Recurrence, and Survival Among Women with Triple-Negative Breast Cancer in the National Comprehensive Cancer Network. *Cancer* 118, 5463–5472. doi:10.1002/cncr.27581

Liu, L. L., Zhao, H., Ma, T. F., Ge, F., Chen, C. S., and Zhang, Y. P. (2015). Identification of Valid Reference Genes for the Normalization of RT-qPCR Expression Studies in Human Breast Cancer Cell Lines Treated with and without Transient Transfection. *PLoS One* 10, e0117058. doi:10.1371/journal.pone.0117058

Man, S. M., and Kanneganti, T.-D. (2015). Regulation of Inflammasome Activation. *Immunol. Rev.* 265, 6–21. doi:10.1111/imr.12296

Nolan, E., Savas, P., Policheni, A. N., Darcy, P. K., Vaillant, F., Mintoff, C. P., et al. (2017). Combined Immune Checkpoint Blockade as a Therapeutic Strategy for BRCA1-Mutated Breast Cancer. *Sci. Transl Med.* 9, eaal4922. doi:10.1126/scitranslmed.aal4922

Orning, P., Weng, D., Starheim, K., Ratner, D., Best, Z., Lee, B., et al. (2018). Pathogen Blockade of TAK1 Triggers Caspase-8-Dependent Cleavage of Gasdermin D and Cell Death. *Science* 362, 1064–1069. doi:10.1126/science.aau2818

Park, Y. H., Lee, S. J., Cho, E. Y., La. Choi, Y., Lee, J. E., Nam, S. J., et al. (2019). Clinical Relevance of TNM Staging System According to Breast Cancer Subtypes. *Ann. Oncol.* 30, 2011. doi:10.1093/annonc/mdz223

Pizato, N., Luzete, B. C., Kiffer, L. F. M. V., Corrêa, L. H., de Oliveira Santos, I., Assumpção, J. A. F., et al. (2018). Omega-3 Docosahexaenoic Acid Induces Pyroptosis Cell Death in Triple-Negative Breast Cancer Cells. *Sci. Rep.* 8, 1952. doi:10.1038/s41598-018-20422-0

Pohl, S.-G., Brook, N., Agostino, M., Arfuso, F., Kumar, A. P., and Dharmarajan, A. (2017). Wnt Signaling in Triple-Negative Breast Cancer. *Oncogenesis* 6, e310. doi:10.1038/oncsis.2017.14

Pontén, F., Jirström, K., and Uhlen, M. (2008). The Human Protein Atlas-A Tool for Pathology. *J. Pathol.* 216, 387–393. doi:10.1002/path.2440

Provance, O. K., Geanes, E. S., Lui, A. J., Roy, A., Holloran, S. M., Gunewardena, S., et al. (2021). Disrupting Interferon-Alpha and NF-kappaB Crosstalk Suppresses IFITM1 Expression Attenuating Triple-Negative Breast Cancer Progression. *Cancer Lett.* 514, 12–29. doi:10.1016/j.canlet.2021.05.006

Ranstam, J., and Cook, J. A. (2018). LASSO Regression [J]. *Br. J. Surg.* 105 (10), 1348. doi:10.1002/bjs.10895

Reisenbichler, E. S., Han, G., Bellizzi, A., Bossuyt, V., Brock, J., Cole, K., et al. (2020). Prospective Multi-Institutional Evaluation of Pathologist Assessment of PD-L1 Assays for Patient Selection in Triple Negative Breast Cancer. *Mod. Pathol.* 33, 1746–1752. doi:10.1038/s41379-020-0544-x

Schmid, P., Adams, S., Rugo, H. S., Schneeweiss, A., Barrios, C. H., Iwata, H., et al. (2018). Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer. *N. Engl. J. Med.* 379, 2108–2121. doi:10.1056/NEJMoa1809615

Shao, W., Yang, Z., Fu, Y., Zheng, L., Liu, F., Chai, L., et al. (2021). The Pyroptosis-Related Signature Predicts Prognosis and Indicates Immune Microenvironment Infiltration in Gastric Cancer. *Front. Cel Dev. Biol.* 9, 676485. doi:10.3389/fcell.2021.676485

Shi, J., Zhao, Y., Wang, K., Shi, X., Wang, Y., Huang, H., et al. (2015). Cleavage of GSDMD by Inflammatory Caspases Determines Pyroptotic Cell Death. *Nature* 526, 660–665. doi:10.1038/nature15514

Smyth, G. K. (2005). *Limma: Linear Models for Microarray Data, Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer, 397–420.

Sullivan, L. M., Massaro, J. M., and D'Agostino, R. B. (2004). Presentation of Multivariate Data for Clinical Use: The Framingham Study Risk Score Functions. *Statist. Med.* 23, 1631–1660. doi:10.1002/sim.1742

Therneau, T. M., and Lumley, T. (2015). Package 'survival' [J]. *R. Top. Doc.* 128 (10), 28–33.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al. (2014). The Dynamics and Regulators of Cell Fate Decisions Are Revealed by Pseudotemporal Ordering of Single Cells. *Nat. Biotechnol.* 32, 381–386. doi:10.1038/nbt.2859

Vikas, P., Borcherding, N., and Zhang, W. (2018). The Clinical Promise of Immunotherapy in Triple-Negative Breast Cancer. *Cmar* 10, 6823–6833. doi:10.2147/CMAR.S185176

Wang, Z.-Q., Milne, K., Derocher, H., Webb, J. R., Nelson, B. H., and Watson, P. H. (2017). PD-L1 and Intratumoral Immune Response in Breast Cancer. *Oncotarget* 8, 51641–51651. doi:10.18632/oncotarget.18305

Whitehead, M. J., McCanney, G. A., Willison, H. J., and Barnett, S. C. (2019). MyelinJ: an ImageJ Macro for High Throughput Analysis of Myelinating Cultures. *Bioinformatics* 35, 4528–4530. doi:10.1093/bioinformatics/btz403

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26, 1572–1573. doi:10.1093/bioinformatics/btq170

Wu, J., Cai, Y., Zhao, G., and Li, M. (2021). A Ten N6-Methyladenosine-Related Long Non-Coding RNAs Signature Predicts Prognosis of Triple-Negative Breast Cancer. *J. Clin. Lab. Anal.* 35, e23779. doi:10.1002/jcla.23779

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a Resource for Therapeutic Biomarker Discovery in Cancer Cells. *Nucleic Acids Res.* 41, D955–D961. doi:10.1093/nar/gks1111

Yang, F., Cao, L., Sun, Z., Jin, J., Fang, H., Zhang, W., et al. (2016). Evaluation of Breast Cancer Stem Cells and Intratumor Stemness Heterogeneity in Triple-Negative Breast Cancer as Prognostic Factors. *Int. J. Biol. Sci.* 12, 1568–1577. doi:10.7150/ijbs.16874

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

Zaki, M. H., Vogel, P., Body-Malapel, M., Lamkanfi, M., and Kanneganti, T.-D. (2010). IL-18 Production Downstream of the Nlrp3 Inflammasome Confers protection against Colorectal Tumor Formation. *J.I.* 185, 4912–4920. doi:10.4049/jimmunol.1002046

Zanker, D. J., Spurling, A. J., Brockwell, N. K., Owen, K. L., Zakhour, J. M., Robinson, T., et al. (2020). Intratumoral Administration of the Toll-like Receptor 7/8 Agonist 3M-052 Enhances Interferon-Driven Tumor Immunogenicity and Suppresses Metastatic Spread in Preclinical Triple-Negative Breast Cancer. *Clin. Transl Immunol.* 9, e1177. doi:10.1002/cti2.1177

Zhang, Z., Zhang, Y., Xia, S., Kong, Q., Li, S., Liu, X., et al. (2020). Gasdermin E Suppresses Tumour Growth by Activating Anti-tumour Immunity. *Nature* 579, 415–420. doi:10.1038/s41586-020-2071-9

Zhou, Z., He, H., Wang, K., Shi, X., Wang, Y., Su, Y., et al. (2020). Granzyme A from Cytotoxic Lymphocytes Cleaves GSDMB to Trigger Pyroptosis in Target Cells. *Science* 368, eaaz7548. doi:10.1126/science.aaz7548

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.

GLOSSARY

AIM2 absent in melanoma 2
BC breast cancer
CASP1/3/4/5/8 caspase1/3/4/5/8
CDF cumulative distribution function
CR complete remission
DEGs differentially expressed genes
DEmiRNAs differentially expressed miRNAs
EPSTI1 epithelial-stromal interaction 1
ER estrogen receptor
FDA Food and Drug Administration
GDSC Genomics of Drug Sensitivity in Cancer
GEO Gene Expression Omnibus
GSDMB gasdermin B
GSDMD gasdermin D
GSDME gasdermin E
GSVA gene set variation analysis
GTEx Genotype-Tissue Expression
GZMA granzyme A
GZMB granzyme B
HER2 human epidermal growth factor receptor 2
HPA Human Protein Atlas
IC50 half-maximal inhibitory concentration
ICI immune cell infiltration
IFITM1 interferon inducible transmembrane 1
IHC immunohistochemistry
KEGG Kyoto Encyclopedia of Genes and Genomes
mUC metastatic urothelial carcinoma
NLR nucleotide-binding domain and leucine-rich repeat receptor
OS overall survival
PCA principal component analysis
PCD programmed cell death
PD programmed cell death
PD-L1 PD-ligand 1
PR progesterone receptor
scRNA-seq single-cell RNA sequencing
TCGA The Cancer Genome Atlas
TME tumor microenvironment
TNBC triple-negative breast cancer
Tregs regulatory T cells
tSNE t-distributed statistical neighbor embedding.

frontiers | Frontiers in Genetics

METHODS
published: 01 April 2022
doi: 10.3389/fgene.2022.802838

Decosus: An R Framework for Universal Integration of Cell Proportion Estimation Methods

Chinedu A. Anene[1,2]*, *Emma Taggart*[3], *Catherine A. Harwood*[4,5], *Daniel J. Pennington*[3] *and Jun Wang*[1]

[1]Centre for Cancer Genomics and Computational Biology, Barts Cancer Institute, Queen Mary University of London, London, United Kingdom, [2]Centre for Cancer Biology and Therapy, School of Applied Science, London South Bank University, London, United Kingdom, [3]Centre for Immunobiology, Barts and the London School of Medicine, Blizard Institute, Queen Mary University of London, London, United Kingdom, [4]Centre for Cell Biology and Cutaneous Research, Barts and The London School of Medicine and Dentistry, Blizard Institute, Queen Mary University of London, London, United Kingdom, [5]Department of Dermatology, The Royal London Hospital, Barts Health NHS Trust, London, United Kingdom

The assessment of the cellular heterogeneity and abundance in bulk tissue samples is essential for characterising cellular and organismal states. Computational approaches to estimate cellular abundance from bulk RNA-Seq datasets have variable performances, often requiring benchmarking matrices to select the best performing methods for individual studies. However, such benchmarking investigations are difficult to perform and assess in typical applications because of the absence of gold standard/ground-truth cellular measurements. Here we describe Decosus, an R package that integrates seven methods and signatures for deconvoluting cell types from gene expression profiles (GEP). Benchmark analysis on a range of datasets with ground-truth measurements revealed that our integrated estimates consistently exhibited stable performances across datasets than individual methods and signatures. We further applied Decosus to characterise the immune compartment of skin samples in different settings, confirming the well-established Th1 and Th2 polarisation in psoriasis and atopic dermatitis, respectively. Secondly, we revealed immune system-related UV-induced changes in sun-exposed skin. Furthermore, a significant motivation in the design of Decosus is flexibility and the ability for the user to include new gene signatures, algorithms, and integration methods at run time.

Keywords: cell deconvolution, R package, method integration, gene expression, immuno-biology

Edited by:
Luis Zapata,
Institute of Cancer Research (ICR), United Kingdom

Reviewed by:
Florent Petitprez,
University of Edinburgh, United Kingdom
Giulio Caravagna,
University of Trieste, Italy

****Correspondence:***
Chinedu A. Anene
a.anene@qmul.ac.uk

Specialty section:
This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: *27 October 2021*
Accepted: *04 March 2022*
Published: *01 April 2022*

Citation:
Anene CA, Taggart E, Harwood CA, Pennington DJ and Wang J (2022) Decosus: An R Framework for Universal Integration of Cell Proportion Estimation Methods. Front. Genet. 13:802838. doi: 10.3389/fgene.2022.802838

INTRODUCTION

Gene expression quantification is indispensable for the interrogation of cellular and organismal states. However, bulk tissue samples of interest in clinical research have considerable cellular heterogeneity that standard methods (Microarray and RNA-Seq) cannot decipher. Although single-cell technologies have been developed to uncover the cellular heterogeneity within cell populations, they have a range of limitations (e.g., time, tissue types, dropouts, technical noise, and cost), making large-scale or clinical applications impossible. Since bulk analysis methods report average expression levels, it is often challenging to disentangle changes in cell-type composition from fundamental differences in states. To address this, several computational tools (so-called deconvolution methods) have been developed to estimate cell-type composition within bulk expression data. These utilise various models, including least squares regression (Abbas et al., 2009), constrained least squares regression (Li et al., 2016; Racle et al., 2017; Finotello and Trajanoski, 2018), quadratic programming (Gong et al., 2011; Gong and Szustakowski, 2013; Zhong et al., 2013), support vector regression (Newman et al., 2015), the geometric mean of marker gene expression (Becht et al., 2016) and single-sample gene set enrichment analysis (ssGSEA) (Aran et al., 2017), extensively reviewed in (Sturm et al., 2019).

Unfortunately, individual tools have limitations that affect their effective utilisation. Critically, due to the differences in the underlying statistical assumptions and marker gene signatures, the approaches often produce different results in a data type-dependent manner (Jiménez-Sánchez et al., 2019; Sturm et al., 2019). Thus, there is a need for a consensus approach that can combine and integrate these methods into a single robust output.

Recent benchmarking studies (Jiménez-Sánchez et al., 2019; Sturm et al., 2019) have provided robust frameworks to systematically integrate the outputs of these methods into a single unified deconvolution solution, thereby reducing the limitations of individual methods. While these approaches are helpful, significant challenges remain. Jiménez-Sánchez et al., restricted their implementation to a set of common cancers (Jiménez-Sánchez et al., 2019), while Sturm et al., focused on method performance on key immune cell types (Sturm et al., 2019). Thus, neither tool can be applied universally, especially for poorly studied cancers, and non-cancer conditions or deconvolution of non-immune cells. We developed a deconvolution integration tool (Decosus, available as an R package) without tissue type restrictions, allowing for universal application of the integration framework. A further advantage of Decosus is the ability to generate both relative (allowing sample-to-sample comparison only) and absolute (also allowing within-sample comparison of one cell type to another) estimations where possible. We demonstrate the utility of Decosus through reference datasets and use cases, including cancer and non-cancer datasets. Decosus generates a validated and robust decomposition of cell composition in various tissue types while allowing for flexible integration of new methods and signatures as they become available (github.com/caanene1/Decosus).

MATERIALS AND METHODS

Deconvolution Method

We used common cell deconvolution methods and signatures to estimate cell composition from bulk RNA-Seq datasets individually. The methods include EPIC (Racle et al., 2017), MCPcounter (Becht et al., 2016), quanTISeq (Finotello et al., 2019), and xCell (Aran et al., 2017). The estimates for the four methods were generated using the corresponding R implementations: MCPcounter (v1.2.0), EPIC (v1.1.5), xCell (v1.1.0), and quanTISeq (github.com/icbi-lab/quanTIseq). In addition to these, we used the following gene signatures: Danaher, Davoli, Rooney (Bindea et al., 2013; Danaher et al., 2017; Davoli et al., 2017), and averaged the gene expressions to provide measures of cellular abundance for each signature. Additionally, we provide optional gene signatures from the CellMarker database (Zhang et al., 2019) and use this in the analyses featured here.

TABLE 1 | Default deconvolution method and signatures included in Decosus.

Signature/Method	Type	Comparison
Xcell	Algorithm	Across Samples
MCP-counter	Algorithm	Across Samples
quanTISeq	Algorithm	Across Samples & Between Cells
EPIC	Algorithm	Across Samples & Between Cells
Danaher	Signature	Across Samples
Davoli	Signature	Across Samples
Rooney	Signature	Across Samples

Decosus Framework

Decosus is an R package that flexibly integrates the estimates of cell compositions from deconvolution methods and cell signatures. We identified cell types for which estimates exist in at least two of the seven default methods and signatures (**Supplementary S1 CSV**). This mapping was similar to that of a previously published method, Consensus$^{\mathrm{TME}}$, where cells were exhaustively mapped to a controlled vocabulary of cell types (Jiménez-Sánchez et al., 2019). However, in Decosus, the user can expand the consensus through the optional arguments for new signatures and mappings (https://github.com/caanene1/Decosus), ensuring that Decosus can produce results relevant to the most up-to-date cell signatures and allowing the addition of rarer cell types. For instance, licensing restrictions did not allow us to include CIBERSORT into the package, but a user can still integrate CIBERSORT if you obtain the output or source code from the author's website.

After generating the estimates for the individual methods, we average the values for the source cells to create a single estimate for the cell types in the controlled vocabulary (**Supplementary S1 CSV**). The assumption behind this approach is that it limits any one method from dominating the estimates, thereby ensuring the consensus is closer to the best measure or is the best performing measure (see results section). We also included optional arguments in the function call to specify whether the data is to be scaled or not and the aggregation mode (i.e., mean (default) or geometric mean). Furthermore, two outputs are provided; 1) relative, the default using all methods, and 2) absolute, which is derived by limiting the methods used to those that can be considered absolute cell compositions (EPIC and quanTISeq, reviewed previously by Sturm et al., 2019). This option offers a less comprehensive selection of consensus cells than the relative output but may be helpful in analyses requiring cell-cell comparison, which is not permitted when using all seven algorithms due to the methods involved (see **Table 1** and Sturm et al., 2019). The Decosus R package can be obtained through GitHub (https://github.com/caanene1/Decosus).

Datasets for Benchmarking

To benchmark the Decosus framework, we obtained pre-processed bulk RNA-seq and FACS data from Sturm et al. (2019) through their Github repository (https://github.com/icbi-lab/immune_deconvolution_benchmark/releases/download/v1.0.0-rcl/data.tar.gz). This data includes eight healthy PBMC samples (Hoek data) (Finotello et al., 2019), four metastatic melanoma samples (Racle data) (Racle et al., 2017) and three ovarian cancer ascites samples (Schelker data, x2 replicates) (Schelker et al., 2017). We obtained 20 PBMC samples analysed by microarray (gene-level values) and flow-cytometry from the CIBERSORT web portal (Cibersort data) (Newman et al., 2015). Additionally, we obtained the two (SDY311, *n* = 76; SDY420 ref, and *n* = 105) pre-processed bulk RNA-seq and FACS data from the Immport study (Alpert et al., 2019) through the xCell Github repository (https://github.com/dviraran/xCell/tree/master/vignettes).

To evaluate the performance of Decosus in real-world clinical contexts, we interrogated two additional RNA-Seq datasets. The first dataset contained skin samples of atopic dermatitis (*n* = 54) and psoriasis (*n* = 55) (GSE121212) (Tsoi et al., 2019). The second dataset was normal skin samples (sun-exposed and non-sun exposed) samples from the GTEX project (*n* = 1879) (Lonsdale et al., 2013).

All RNA-Seq gene expression values are expressed as transcript per million (TPM).

Datasets for Simulation Studies

To evaluate how well Decosus estimates known cell compositions compared to individual methods, we simulated cell mixtures from single-cell RNA-Seq datasets. Specifically, we extracted the RNA expression values for select samples from the Panglaodb data portal (Franzén et al., 2019), including SRA701877, SRS3279685, SRA713577, SRS3363004, SRA716608, SRS3391633, SRA779509, SRS3805246, SRA878024, and SRS4660846.

Next, we generated expression profiles as below:

1 Given a dataset of annotated n single cell types, assign random fractions to each cell type (the fractions sum to 100 and integers) (data 1).
2 Generate an expression matrix of single cells with 100 columns by randomly selecting the corresponding fraction of the available samples for the selected cell types to be included in the matrix. Here, we used a random selection to introduce noise like the variation in real datasets (data 2). We allow sampling with replacement if the cell type-specific fraction is bigger than the available single cells.
3 Finally, generate a simulated expression profile by adding the expression values across the rows of data 2 and use data 1 as the ground truth.

We repeat the process 500 times (per data source) with different fractions, samples, and cell types.

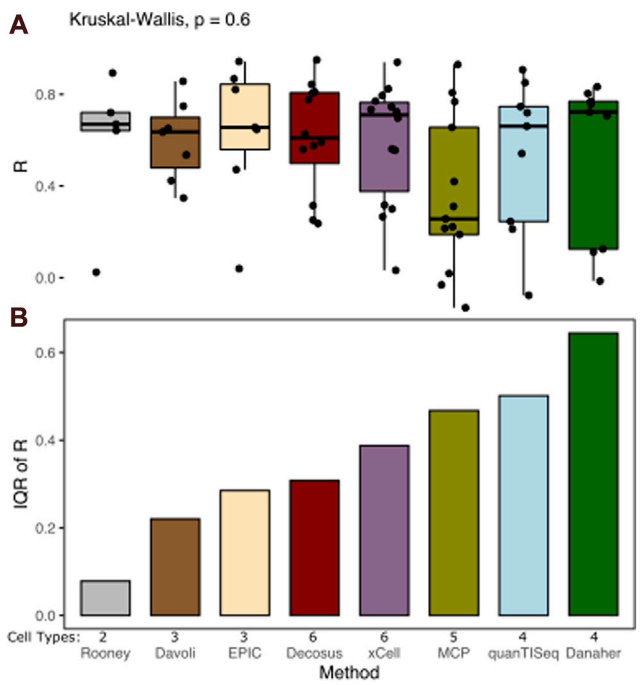


FIGURE 1 | Decosus produces stable estimates of cell proportions. **(A)** Box plot of the Pearson's R of correlation between decomposed and simulated cell fractions (*n* = 500 per single-cell RNA-Seq data used in the simulations studies). The higher the R, the more similar the decomposed and simulated cell fractions are. **(B)** Bar plot of inter-method variability in Pearson's R of correlation between decomposed and simulated cell fractions measured by IQR. The lower the IQR, the more stable the estimates from the method across datasets and cell types. Within the plots, colours represent the method of estimating cell fractions from bulk RNA-Seq datasets. The number of cell types covered for each method in the simulation is indicated (Cell Types). MCP = MCPcounter.

RESULTS

Overview of the Decosus Integration Framework

Our framework integrates seven deconvolution algorithms and cell signatures into consensus estimates of cell composition (**Table 1**). We do this by selecting and averaging the shared cell types across the tools (**Supplementary S1 CSV**). It requires a gene expression dataset and a set of optional parameters under a single R function (set to reasonable defaults, see github. com/caanene1/Decosus). The R implementation of the framework allows for flexible inclusion of new algorithms or signatures at run time (see https://github.com/caanene1/Decosus). When available, the final output has two tables representing relative and absolute consensus estimates (see *Materials and Methods*).

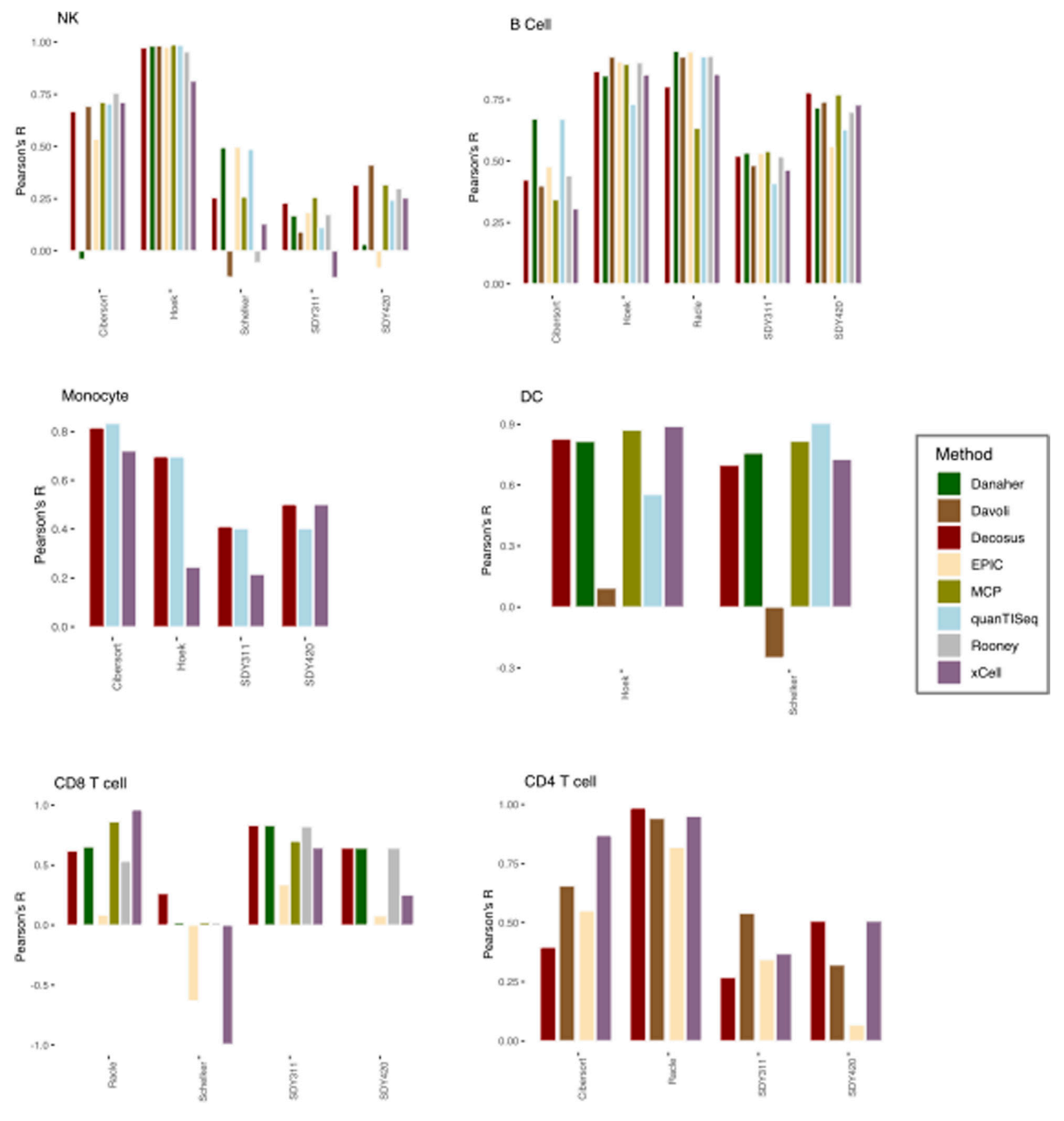


FIGURE 2 | Decosus identifies known cell fractions. Bar plots of the Pearson's R of the correlation between computer-generated cell fractions and flow cytometry fractions. The higher the R, the more similar is the computational estimates and the gold standard. Within the plots colours represent the method of estimating cell fractions from bulk RNA-Seq datasets. MCP = MCPcounter.

Analysis of Simulated Cell Proportions Demonstrate the Stability of Decosus

Cell proportion deconvolution methods produce different results in a data type-dependent manner. Decosus combines and integrates these methods into a single stable consensus value to reduce the data-dependent differences in performance. To assess how well the framework achieves this aim, we first evaluated its ability to decompose known cell proportions using simulated gene expression profiles (see Methods). We simulated large sets of bulk-expression profiles (n = 2,500) with specific cell proportions from multiple single-cell RNA-Seq datasets (source, n = 5). We used multiple sources to reflect data-specific differences in real applications. Interrogating the correlation coefficient between the estimates and the expected cell proportions across a range of cell types revealed surprisingly stable estimates of cell proportions (**Figure 1**). Decosus had the 4th lowest interquartile range (IQR = 0.31) compared to xCell (IQR = 0.39), MCPcounter (IQR = 0.47), quanTISeq (IQR = 0.50), and Danaher (IQR = 0.64) (**Figure 1A**). Although Rooney (IQR = 0.09), Davoli (IQR = 0.22), and EPIC (IQR = 0.29) were lower than Decosus, this is potentially due to the small number of cell types they covered in the simulation (Davoli = 3, EPIC = 3, and Rooney = 2), compared to the other methods (Decosus = 6, xCell = 6, MCPcounter = 5, quanTISeq = 4, and Danaher = 4). Indeed, normalising for the number of cell types revealed Decosus has the second-lowest IQR (**Supplementary Figure S1**). The stability of Decosus estimates is due to the robustness of averaging multiple signatures and methods (see Methods). As expected, there is no difference between median R values across the methods (Kruskal–Wallis, $p = 0.6$, **Figure 1B**), suggesting that Decosus increases the stability of the estimates without reducing the average performance expected from individual tools.

Benchmark on Known Cell Proportions Demonstrate the Utility of Decosus

We applied the Decosus framework to six benchmark datasets (see Methods) and compared the estimated proportions with their corresponding flow cytometry fractions used as ground-truth. We also interrogated the distribution of the performance statistic across the individual methods and datasets. To ensure we can perform meaningful comparative analysis, we restricted our analysis to six cell types (natural killer cells, dendritic cells, monocytes, $CD4^+$ T-cells, $CD8^+$ T-cells and B-cells) present in two or more ground truth datasets. Our consensus estimates obtained a high correlation with the flow cytometry fractions (Median R = 0.64) across all datasets and the evaluated cell types (**Figure 2**). This observation was consistent with the contributing methods/signatures being highly concordant with the corresponding gold standards (**Figure 2**). However, we observed unpredictable performance differences across the same cell type for the individual methods (**Figure 2**). For natural killer (NK) cells, Danaher (r = −0.04038) and EPIC (r = 0.53) had the worst performances in the Cibersort dataset, Rooney (r = −0.0557), xCell (r = 0.129) and Davoli (r = −0.125) in Schelker, and EPIC (r = −0.08), Danaher (r = 0.0316), quanTISeq (r = 0.243), and xCell (r = 0.253) in SDY420, while all the methods performed well in Hoek or poorly in SDY311 (**Figure 2**). We observed similar behaviour for monocytes, where xCell performed worse in Hoek (r = 0.244) and xCell (r = 0.215) in SDY311. Interestingly, for dendritic cells, Davoli performed worse in both Hoek (r = 0.092) and Schelker (r = 0.25), but QuanTISeq additionally performed poorly in Hoek datasets (r = 0.55). Although T and B cells have multiple subtypes making benchmark analysis difficult, we generally made similar observations for $CD4^+$ T cells, $CD8^+$ T cells and B cells (**Figure 2**), where the different combinations of methods performed worse in different datasets. These observations suggest that no single method can guarantee top performance across user cases, even for the same cell type. However, our consensus estimates consistently performed well across the datasets, regardless of cell types and user cases (**Figure 2**). For instance, it had high performance in all datasets with DC (Hoek, r = 0.82; Schelker, r = 0.69) and most datasets with NK cells (Cibersort, r = 0.668; Hoek, r = 0.973; SDY420, r = 0.315). Furthermore, Decosus was able to derive the cellular estimates for all six cell types benchmarked here, whilst other methods/frameworks were restricted to a subset of cell types where marker gene signatures were included internally (**Figure 2**).

Decosus Identified the Immunological Differences Between Atopic Dermatitis and Psoriasis

To further demonstrate the utility of Decosus and evaluate its performance in other physiological and health settings, we focused on diseases and cell types that were not well covered and annotated by existing methods. Here, we simulated a condition with known condition-dependent differences in cell proportions. Specifically, atopic dermatitis (AD) and psoriasis (PSO) are common skin conditions associated with barrier dysfunction. Both are characterised by T-cell driven inflammation; however, in AD, $CD4^+$ T helper cells (Th) are polarised towards a Th2 phenotype, while Th1 polarization is characteristic of PSO (Brunner et al., 2017; Albanesi, 2019). Thus, immune cell composition estimated from bulk RNA expression profiles from PSO and AD skin samples should enrich for Th1 and Th2 cell signatures, respectively. Note that many existing methods do not have these two cell types, thus could not handle such cases. To this end, we collected 54 AD and 55 PSO samples from a publicly available dataset (GSE121212) (Tsoi et al., 2019) and applied our framework to interrogate the enriched immune profiles. We observed a significantly higher Th1 cell signature in PSO samples than AD samples and the reverse for Th2 (**Figure 3**), aligning with what is widely recognised in the literature (Brunner et al., 2017; Albanesi, 2019). The results also showed expansions in other cell types known to infiltrate each lesion, such as basophils in AD (Mashiko et al., 2017), macrophages and neutrophils in PSO (Lowes et al., 2014) (**Figure 3**). These results indicate that our framework can provide robust estimates of cell proportions in non-cancer bulk tissue samples, like precursor lesions.

Decosus Enables the Identification of UV Mediated Immune Reprogramming in the Skin

Finally, we utilised transcriptomic data from the GTEX project of non-sun exposed and sun-exposed skin to evaluate the

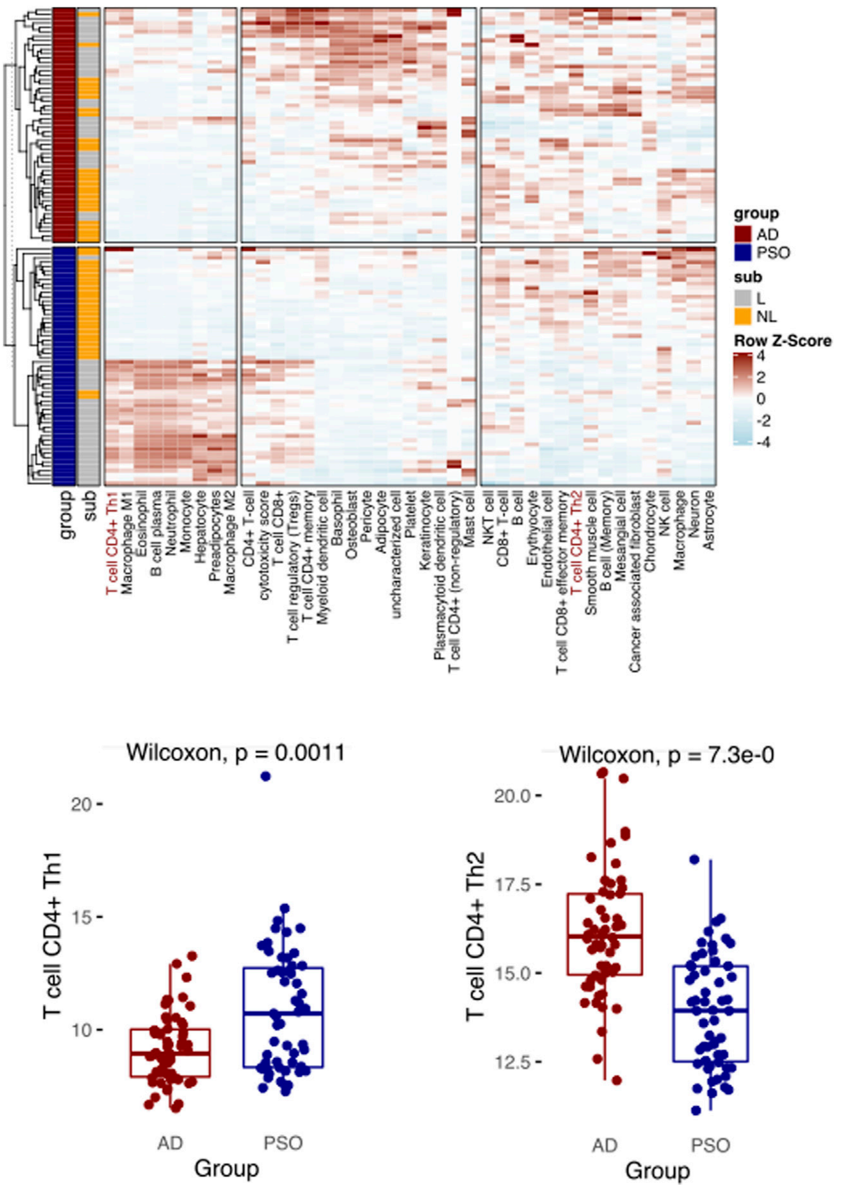


FIGURE 3 | Decosus identifies the Th1 and Th2 differences between atopic dermatitis and psoriasis. (Top) Heatmap of Decosus estimated cell fractions across atopic dermatitis (AD, *n* = 54) and psoriasis (PSO, *n* = 55) samples from GSE121212. Within the plot, Th1 and Th2 cells are indicated in red, while other cells of interest are highlighted with blue. (Bottom) Box plots of the Decosus estimates of Th1 cells (left) and Th2 (right) fractions across skin samples of AD and PSO. Multiple test adjusted T. test *p* values are indicated. Within the heatmap, NL = non-lesional and L = lesional.

impact of ultraviolet radiation (UV) on skin immune profiles. We assessed the difference between the immune cell estimates of the sample groups and visualised the fold change for significant (T-test, $p < 0.05$) cell types. We found that UV exposure significantly enriched several immune cell types in the skin, including monocytes, dendritic cells, and macrophages (**Figure 4**). Interestingly, $CD4^+$ and $CD8^+$ T cells were depleted in sun-exposed skin (**Figure 4**). These observations are consistent with previous studies showing that UV exposure inhibits the expansion of these T cell subtypes while increasing innate immune cells (Rana et al., 2008; Gläser et al., 2009). Reprogramming of T cell composition is consistent with the idea that impaired immune function through UV damage plays a role in skin cancers (Freeman et al., 2014; Slater and Googe, 2016). Indeed, active research programmes in our group are using Decosus to help characterise the immunological factors underlying the progression of actinic keratoses (sun-damaged skin, pre-malignant lesions) to squamous cell carcinomas.

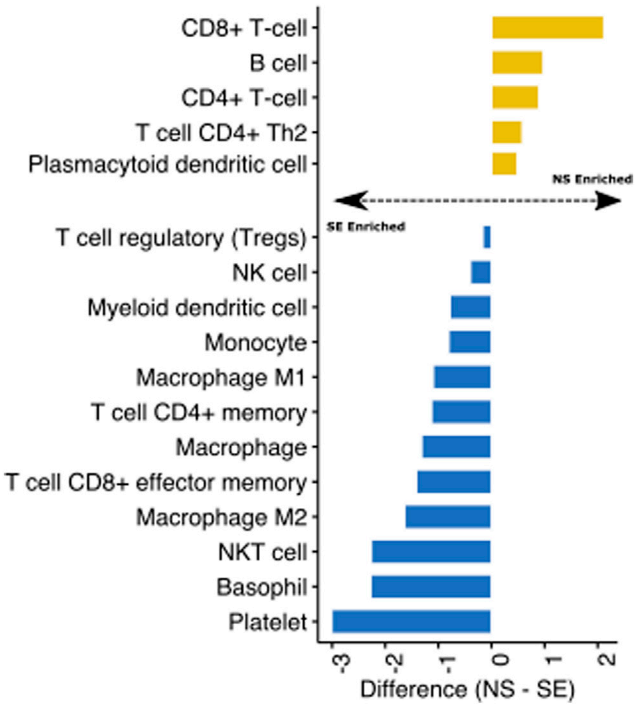


FIGURE 4 | Identification of UV-mediated reprogramming of skin immune cell profile. Bar plot depicting the mean difference between the Decosus immune cell estimates for non-sun exposed skin (NS) and sun-exposed skin (SE) (n = 1879). The plot shows only immune cells significantly different between the two groups at *p*-value < 0.05.

DISCUSSION

Estimating the cellular compositions in bulk RNA-Seq samples has been addressed by various algorithms and gene signatures. These approaches have different assumptions and strengths, frequently resulting in varied performances across the different dataset (Sturm et al., 2019). Thus, it is challenging to objectively select the best method in real applications which lack ground truth cell fractions. The Decosus framework addressed this gap by providing consensus estimates that exhibit consistent performance across different benchmark datasets and has further invaluable features not found in existing consensus methods. The main utilities of Decosus are that it 1) leverages the strengths of the individual methods and signatures while minimising their weakness, and 2) provides a broader coverage of cell types.

Further,our framework is inherently expandable, whereby the user can add new methods, signatures, and cell mappings to the default set. We demonstrated this function by including gene sets from the CellMarker database to enable the consensus estimation of Th1 and Th2 cell fractions in AD and PSO samples (**Figure 2**). This flexible approach is critical for robustly estimating cell abundance and fractions across biological states and represents an important advance in the domain compared to previous methods that exclusively focus on common cancers or cell types. Though the expansion of the signatures and mappings may inadvertently incorporate spurious estimates to the consensus, the multiple sources per cell type can better capture the diversity in cellular gene expression profiles across different biological states. Unlike the Consensus^{TME} that generated new consensus gene signatures (Jiménez-Sánchez et al., 2019), we solved this problem by first calculating the cell estimations individually for each method (see *Methods*) before taking the average value for the same cell in each method. Here, a future update may include additional ways to generate consensus estimates such as geometric mean, trimmed and weighted averages, particularly as validation datasets become available. Further advantages to our tool are the ability to apply Decosus to any tissue type or disease compared to previous efforts to create consensus cell composition estimates, which focused on tumour tissues (Jiménez-Sánchez et al., 2019; Sturm et al., 2019). We also incorporated the option to use absolute estimates, which is crucial for applications requiring cell-cell comparisons.

Benchmarking on the PBMC and cancer datasets showed that although each method ranked highly in at least one dataset, we saw highly varied correlations across datasets. This performance issue is expected but impossible to identify in actual use cases because of the differences in the statistical assumptions and gene signatures associated with each method. Our consensus estimates reduce this unpredictable behaviour by averaging out the poorly performing methods. Indeed, we adequately identified the expected Th1 cells enriched in PSO and the Th2 cells in AD. It is worth noting that many existing methods, such as Cibersort, EPIC, and ConsensusTME, could not resolve such a case due to their limited coverage of diseases/conditions or cell types (see **Supplementary S2 CSV**), further highlighting the versatility of our method.

An important limitation of the Decosus framework is that it represents the aggregate performance of the contributing methods. Thus, if they have universally poor performance for a given case, then Decosus will have a corresponding poor performance. For instance, all the approaches, including Decosus, performed poorly in decomposing NK cells and Monocytes from the SDY311 dataset (**Figure 2**). However, Decosus is stable for most cases compared to the individual methods.

To allow for easy incorporation of Decosus into new and existing workflows, we implemented an object-oriented system in R, allowing the user to add, retrieve and evaluate individual methods (https://github.com/caanene1/Decosus). The full output of Decosus provides consensus estimates when available and offers unified interphase for the procedures. Although, Decosus is implemented and valid for human data, the framework can easily be expanded with new methods and signatures, including for other species at run time, as demonstrated in the analysis of Th1 and Th2 cells in skin samples. Moreover, one of the future directions of Decosus is to create a flexible function within our method that allows users to input associated weights of individual methods and additive equations, to facilitate users to infer the most accurate estimates of cellular compositions in their biological settings. However, this will require very large validation datasets to derive weights accurately for the biological setting of interest.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

Conceptualisation, CA, ET, CH, DP and JW; Software, CA and ET; Investigation, CA, ET, CH, DP and JW; Writing—Original Draft, CA and ET; Writing—Review & Editing, CA, ET, CH, DP and JW; Supervision, CH, DP, and JW.

FUNDING

This study was supported by the Barts Charity project grant (reference MRDU0003), awarded to CH, DP, and JW.

ACKNOWLEDGMENTS

The authors acknowledge support from Cancer Research United Kingdom Centre of Excellence Award to Barts Cancer Centre (reference C16420/A18066). JW acknowledges support from the Academy of Medical Sciences Springboard Award (SBF003\1025), and Cancer Research United Kingdom (C355/A26819) and FC AECC and AIRC under the Accelerator Award Program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.802838/full#supplementary-material

Supplementary Figure 1 | Bar plot of inter-method variability in Pearson's R of correlation between decomposed and simulated cell fractions measured by IQR normalised by the number of cell types covered for each method in the simulation. MCP = MCPcounter.

REFERENCES

Abbas, A. R., Wolslegel, K., Seshasayee, D., Modrusan, Z., and Clark, H. F. (2009). Deconvolution of Blood Microarray Data Identifies Cellular Activation Patterns in Systemic Lupus Erythematosus. *PLoS One* 4, e6098. doi:10.1371/journal.pone.0006098

Albanesi, C. (2019). "Immunology of Psoriasis," in *Clinical Immunology* (Elsevier), 871–878. doi:10.1016/b978-0-7020-6896-6.00064-8

Alpert, A., Pickman, Y., Leipold, M., Rosenberg-Hasson, Y., Ji, X., Gaujoux, R., et al. (2019). A Clinically Meaningful Metric of Immune Age Derived from High-Dimensional Longitudinal Monitoring. *Nat. Med.* 25, 487–495. doi:10.1038/s41591-019-0381-y

Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: Digitally Portraying the Tissue Cellular Heterogeneity Landscape. *Genome Biol.* 18, 220–314. doi:10.1186/s13059-017-1349-1

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., et al. (2016). Estimating the Population Abundance of Tissue-Infiltrating Immune and Stromal Cell Populations Using Gene Expression. *Genome Biol.* 17, 218–220. doi:10.1186/s13059-016-1070-5

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity* 39, 782–795. doi:10.1016/j.immuni.2013.10.003

Brunner, P. M., Guttman-Yassky, E., and Leung, D. Y. M. (2017). The Immunology of Atopic Dermatitis and its Reversibility with Broad-Spectrum and Targeted Therapies. *J. Allergy Clin. Immunol.* 139, S65–S76. doi:10.1016/j.jaci.2017.01.011

Danaher, P., Warren, S., Dennis, L., D'Amico, L., White, A., Disis, M. L., et al. (2017). Gene Expression Markers of Tumor Infiltrating Leukocytes. *J. Immunother. Cancer* 5, 18–15. doi:10.1186/s40425-017-0215-8

Davoli, T., Uno, H., Wooten, E. C., and Elledge, S. J. (2017). Tumor Aneuploidy Correlates with Markers of Immune Evasion and with Reduced Response to Immunotherapy. *Science* 355, eaaf8399. doi:10.1126/science.aaf8399

Finotello, F., and Trajanoski, Z. (2018). Quantifying Tumor-Infiltrating Immune Cells from Transcriptomics Data. *Cancer Immunol. Immunother.* 67, 1031–1040. doi:10.1007/s00262-018-2150-z

Finotello, F., Mayer, C., Plattner, C., Laschober, G., Rieder, D., Hackl, H., et al. (2019). Molecular and Pharmacological Modulators of the Tumor Immune Contexture Revealed by Deconvolution of RNA-Seq Data. *Genome Med.* 11, 34–20. doi:10.1186/s13073-019-0638-6

Franzén, O., Gan, L.-M., and Björkegren, J. L. (2019). PanglaoDB: a Web Server for Exploration of Mouse and Human Single-Cell RNA Sequencing Data. *Database* 2019, baz046. doi:10.1093/database/baz046

Freeman, A., Bridge, J. A., Maruthayanar, P., Overgaard, N. H., Jung, J.-W., Simpson, F., et al. (2014). Comparative Immune Phenotypic Analysis of Cutaneous Squamous Cell Carcinoma and Intraepidermal Carcinoma in Immune-Competent Individuals: Proportional Representation of CD8+ T-Cells but Not FoxP3+ Regulatory T-Cells Is Associated with Disease Stage. *PLoS One* 9, e110928. doi:10.1371/journal.pone.0110928

Gläser, R., Navid, F., Schuller, W., Jantschitsch, C., Harder, J., Schröder, J. M., et al. (2009). UV-B Radiation Induces the Expression of Antimicrobial Peptides in Human Keratinocytes *In Vitro* and *In Vivo*. *J. Allergy Clin. Immunol.* 123, 1117–1123. doi:10.1016/j.jaci.2009.01.043

Gong, T., and Szustakowski, J. D. (2013). DeconRNASeq: a Statistical Framework for Deconvolution of Heterogeneous Tissue Samples Based on mRNA-Seq Data. *Bioinformatics* 29, 1083–1085. doi:10.1093/bioinformatics/btt090

Gong, T., Hartmann, N., Kohane, I. S., Brinkmann, V., Staedtler, F., Letzkus, M., et al. (2011). Optimal Deconvolution of Transcriptional Profiling Data Using Quadratic Programming with Application to Complex Clinical Blood Samples. *PLoS One* 6, e27156. doi:10.1371/journal.pone.0027156

Jiménez-Sánchez, A., Cast, O., and Miller, M. L. (2019). Comprehensive Benchmarking and Integration of Tumor Microenvironment Cell Estimation Methods. *Cancer Res.* 79, 6238–6246. doi:10.1158/0008-5472.CAN-18-3560

Li, B., Severson, E., Pignon, J. C., Zhao, H., Li, T., Novak, J., et al. (2016). Comprehensive Analyses of Tumor Immunity: Implications for Cancer Immunotherapy. *Genome Biol.* 17, 174–216. doi:10.1186/s13059-016-1028-7

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., et al. (2013). The Genotype-Tissue Expression (GTEx) Project. *Nat. Genet.* 45, 580–585. doi:10.1089/bio.2015.29031.hmm

Lowes, M. A., Suarez-Farinas, M., and Krueger, J. G. (2014). Immunology of Psoriasis. *Annu. Rev. Immunol.* 32, 227–255. doi:10.1146/annurev-immunol-032713-120225

Mashiko, S., Mehta, H., Bissonnette, R., and Sarfati, M. (2017). Increased Frequencies of Basophils, Type 2 Innate Lymphoid Cells and Th2 Cells in Skin of Patients with Atopic Dermatitis but Not Psoriasis. *J. Dermatol. Sci.* 88, 167–174. doi:10.1016/j.jdermsci.2017.07.003

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods* 12, 453–457. doi:10.1038/nmeth.3337

Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E., and Gfeller, D. (2017). Simultaneous Enumeration of Cancer and Immune Cell Types from Bulk Tumor Gene Expression Data. *elife* 6, e26476. doi:10.7554/eLife.26476

Rana, S., Byrne, S. N., MacDonald, L. J., Chan, C. Y.-Y., and Halliday, G. M. (2008). Ultraviolet B Suppresses Immunity by Inhibiting Effector and Memory T Cells. *Am. J. Pathol.* 172, 993–1004. doi:10.2353/ajpath.2008.070517

Schelker, M., Feau, S., Du, J., Ranu, N., Klipp, E., MacBeath, G., et al. (2017). Estimation of Immune Cell Content in Tumour Tissue Using Single-Cell RNA-Seq Data. *Nat. Commun.* 8, 2032–2112. doi:10.1038/s41467-017-02289-3

Slater, N. A., and Googe, P. B. (2016). PD-L1 Expression in Cutaneous Squamous Cell Carcinoma Correlates with Risk of Metastasis. *J. Cutan. Pathol.* 43, 663–670. doi:10.1111/cup.12728

Sturm, G., Finotello, F., Petitprez, F., Zhang, J. D., Baumbach, J., Fridman, W. H., et al. (2019). Comprehensive Evaluation of Transcriptome-Based Cell-type Quantification Methods for Immuno-Oncology. *Bioinformatics* 35, i436–i445. doi:10.1093/bioinformatics/btz363

Tsoi, L. C., Rodriguez, E., Degenhardt, F., Baurecht, H., Wehkamp, U., Volks, N., et al. (2019). Atopic Dermatitis Is an IL-13-Dominant Disease with Greater Molecular Heterogeneity Compared to Psoriasis. *J. Invest. Dermatol.* 139, 1480–1489. doi:10.1016/j.jid.2018.12.018

Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., Deng, C., et al. (2019). CellMarker: a Manually Curated Resource of Cell Markers in Human and Mouse. *Nucleic Acids Res.* 47, D721–D728. doi:10.1093/nar/gky900

Zhong, Y., Wan, Y. W., Pang, K., Chow, L. M., and Liu, Z. (2013). Digital Sorting of Complex Tissues for Cell Type-specific Gene Expression Profiles. *BMC Bioinformatics* 14, 89–10. doi:10.1186/1471-2105-14-89

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.

ORIGINAL RESEARCH
published: 18 May 2022
doi: 10.3389/fgene.2022.805256

NFE2L3 as a Novel Biomarker Associated With IL-2/STAT5/NLRP3 Signaling Pathway in Malignant Pleural Mesothelioma and Other Cancers

Zhen Wang[†], *Han Yang*[†], *Bin Luo*, *Pengfei Duan and Peng Lin* *

Department of Thoracic Oncology, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, China

Edited by:
Eszter Lakatos,
Queen Mary University of London,
United Kingdom

Reviewed by:
Haitang Yang,
Shanghai Jiao Tong University, China
Akinobu Ota,
Aichi Medical University, Japan

Correspondence:
Peng Lin
linpeng@sysucc.org.cn

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: *30 October 2021*
Accepted: *19 April 2022*
Published: *18 May 2022*

Citation:
Wang Z, Yang H, Luo B, Duan P and Lin P (2022) NFE2L3 as a Novel Biomarker Associated With IL-2/STAT5/NLRP3 Signaling Pathway in Malignant Pleural Mesothelioma and Other Cancers.
Front. Genet. 13:805256.
doi: 10.3389/fgene.2022.805256

Background: Malignant pleural mesothelioma (MPM) is a malignant tumor originating from pleural mesothelial cells and has a high mortality rate worldwide. With the advent of immunotherapy in MPM treatment, there is an urgent need to elucidate the immune-related mechanisms in this caner.

Methods: Single-sample gene set enrichment analysis (ssGSEA) was used to score the immunocytes infiltration of data from different database sources. Identification of immunocyte-related genes was performed with weighted gene co-expression network analysis (WGCNA), differentially expressed genes (DEGs) analysis, and correlation analysis. Pan-caner analysis was performed using “DiffExp” and “Correlation” modules in TIMER.

Results: T-helper 2 (Th2) cell was found to be a poor prognostic factor for patients with MPM. Then a transcription factor, NFE2L3, was identified as a biomarker that showed a strong positive correlation with Th2 cell infiltration, and was highly expressed in MPM tissues and was related to the poor prognosis of these patients. At the same time, multiple NFE2L3 methylation sites were negatively correlated with Th2 cell infiltration, and patients with a high degree of methylation enjoy a better prognosis. Pan-caner analysis indicated that NFE2L3 might promote the differentiation of Th2 cells through the IL-2/STAT5/NLRP3 signaling pathway in MPM and many other cancers.

Conclusion: We believe that NFE2L3 can serve as a potential biomarker related to the diagnosis and prognosis of patients with MPM, and speculate that NFE2L3 could promote Th2 cell differentiation *via* IL-2/STAT5/NLRP3 signaling pathway in MPM and many other cancers.

Keywords: NFE2L3, IL-2, Stat5, NLRP3, t-helper 2 cell, malignant pleural mesothelioma

INTRODUCTION

Malignant pleural mesothelioma (MPM) originates from pleural mesothelial cells and is a relatively rare type of cancer, which accounts for 0.3% of all cancer cases (Fernandez-Cuesta et al., 2021). Due to its aggressiveness and difficulty in early diagnosis, MPM is difficult to be cured. Because of the limited role of surgery in the treatment of MPM, the chemotherapy regimen of pemetrexed combined with platinum has occupied a dominant position in the treatment of MPM for a long time (de Gooijer et al., 2018). However, survival benefit from chemotherapy is limited, and the 5-years survival rate of patients with MPM is still less than 10% (Kindler et al., 2018).

In recent years, immune checkpoint inhibitors (ICIs) have been proven to improve the prognosis of various solid tumors, and their anti-tumor effects in MPM have gradually become clear (Lievense et al., 2017). A phase III clinical trial (Checkmate 743) has proved for the first time that compared with chemotherapy, the first-line treatment of nivolumab combined with ipilimumab can provide significant improvements in overall survival (OS) for patients with advanced MPM (Baas et al., 2021). However, considering the complexity of the tumor immune microenvironment, we still need to explore more immune-related mechanisms and targets to increase our understanding of MPM and treatment methods for this disease.

As an essential part of the human immune system, $CD4^+$ T cells play a pivotal role in adaptive immune responses. Under the activation of external cytokines, naive $CD4^+$ T cells can differentiate into multiple T helper cells, including Th1, Th2, Th17 (Zhu et al., 2010). Among them, Th2 cells have been confirmed to have immunosuppressive effects in many tumors (Kusuda et al., 2005; Nevala et al., 2009; De Monte et al., 2011). In normal tissues, T helper lymphocyte subsets are in equilibrium. However, tumor cells can secrete a variety of cytokines, putting the body in a state where Th2 cells dominate, leading to immune escape and tumor progression. Therefore, in this study, we focused on the effect of Th2 cells in patients with malignant pleural mesothelioma and explored a novel regulatory network of Th2 cell differentiation using a series of bioinformatics analysis methods.

MATERIALS AND METHODS

Analysis Overview

In this study, utilizing transcriptomes downloaded from TCGA and GEO, we firstly identified that Th2 cell was associated with prognosis of patients with MPM, and NFE2L3, a transcription factor (TF), was associated with infiltration of this immunocyte. Then, using DNA methylation data from TCGA, we analyzed the correlation of NFE2L3's methylation site with the infiltration of Th2 cell and the prognosis of patients with MPM. Meanwhile, using tissue sections from patients with MPM, we verified the expression of NFE2L3 in normal pleural tissues and tumor tissues. Finally, with the help of TIMER database, we analyzed the correlation between NFE2L3 and Th2 cell regulatory pathway-related genes in mesothelioma (MESO) and many other cancers. The detailed research process is shown in the flow chart (**Figure 1**).

Date Source

All data used in this study are from public databases, including Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) (Barrett et al., 2013; Cancer Genome Atlas Research et al., 2013). The clinical data, transcriptome data in FPKM format and DNA methylation data of malignant pleural mesothelioma (MPM) in TCGA were downloaded from the National Cancer Institute's (NCI's) Genomic Data Commons (GDC) (https://portal.gdc.cancer.gov/) (Zhang et al., 2021). Search with "mesothelioma" as a keyword, we found and downloaded gene expression profiles of GSE51024 and GSE163720 from GEO database (http://www.ncbi.nlm.nih.gov/geo/). GSE51024 is composed of 55 MPM tissues and 41 normal paired lung parenchyma tissues. And GSE163720 contains 131 tumor samples from patients with MPM.

Single-Sample Gene Set Enrichment Analysis

Immunocyte-related gene sets were got from The Cancer Immunome database (TCIA) (Charoentong et al., 2017). Through the expression value of 782 immune-related genes, we scored infiltration levels of 28 types of immunocytes. Immunocytes infiltration levels of samples were quantified by the ssGSEA algorithm in R package GSVA (Hanzelmann et al., 2013). Through Kaplan-Meier survival analysis, we found out the target immunocyte associated with the prognosis of patients with MPM.

Weighted Gene Co-Expression Network Analysis

The transcriptome data of samples from TCGA were further analyzed utilizing WGCNA to find a gene set highly correlated with infiltration level of the target immunocyte. As an algorithm for transcriptome analysis, WGCNA can identify genes with highly correlated expression patterns, and calculate the correlation between the gene set and clinicopathological traits of samples (Langfelder and Horvath, 2008).

R package "WGCNA" was used to complete the calculation process in this step. First, the 5,000 genes with the highest average expression were selected for the subsequent analysis. Then, the optimal soft-thresholding power for network construction was calculated, and module eigengenes (MEs) containing a series of co-expressed genes were constructed using a dynamic tree-cutting algorithm. Finally, the co-expressed gene set with the strongest correlation with Th2 cell infiltration can be found by analyzing the correlation between MEs and clinicopathological traits.

Functional Enrichment Analysis

Metascape, a meta-analysis website, was used for functional enrichment analysis of the co-expressed gene set (Zhou et al., 2019). By using the "Express Analysis" module in Metascape, the enriched biological processes and pathways of the selected co-expressed gene set were obtained.

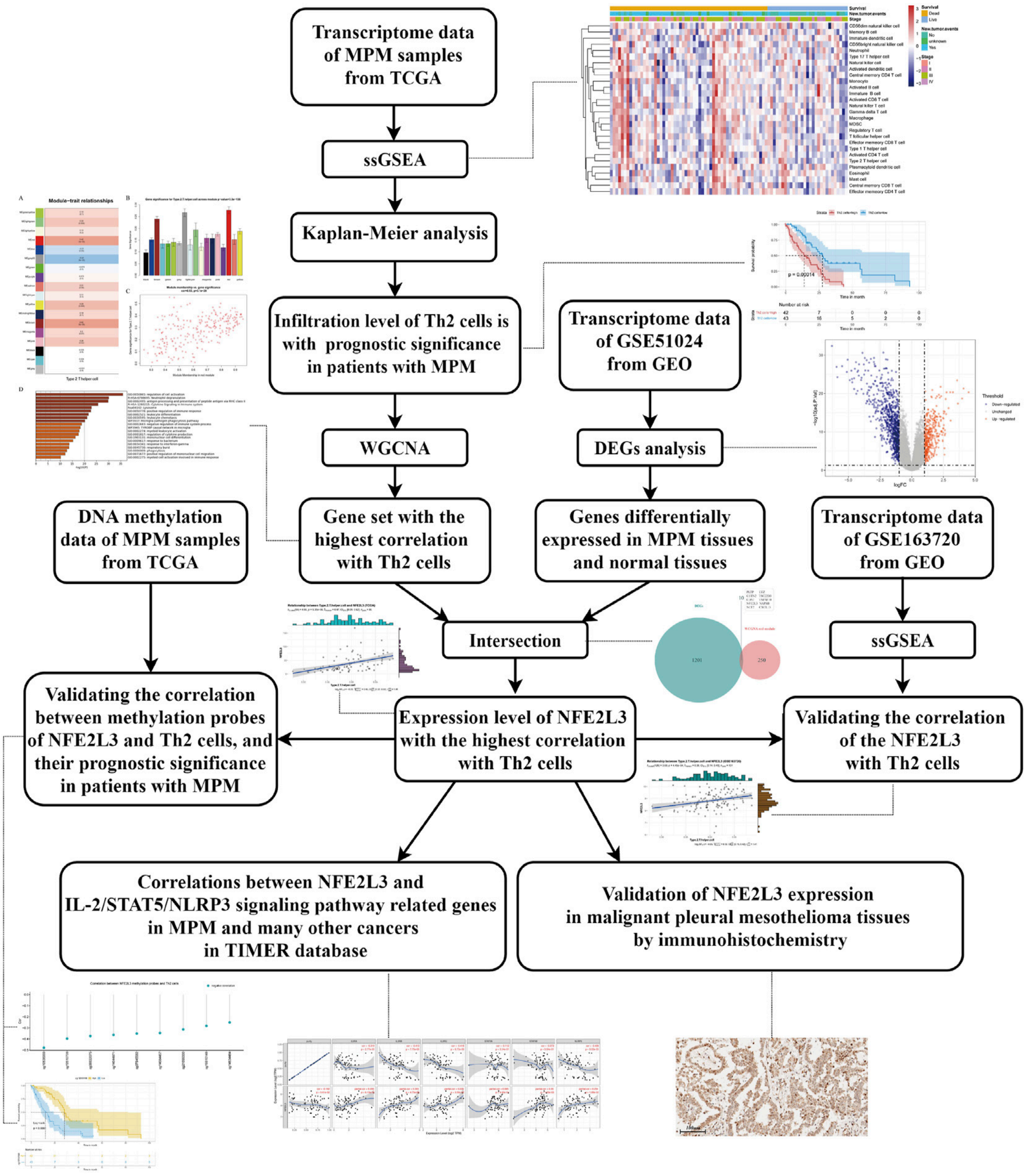


FIGURE 1 | The detailed workflow of this study. MPM, malignant pleural mesothelioma; TCGA, The Cancer Genome Atlas; ssGSEA, single-sample gene set enrichment analysis; Th2 cells, type 2 T helper cells; WGCNA, weighted gene co-expression network analysis; GEO, Gene Expression Omnibus; DEGs, differentially expressed genes.

Differentially Expressed Genes and Correlation Analysis

GSE51024 is composed of 96 samples, including 55 MPM tissues and 41 normal paired lung parenchyma tissues. Using GSE51024 for DEGs analysis, we can screen out genes that are significantly up-regulated or down-regulated in MPM tissues compared to normal lung parenchyma tissues. DEGs were analyzed using "limma" package in R, with an adjusted *p* value < 0.05 and |logFC| >1 (Ritchie et al., 2015). Mann-Whitney test was performed to calculate expression differences between tumor and normal tissues (Perme and Manevski, 2019).

The intersection of DEGs and selected MEs was used to further analyze the correlation with infiltration level of the target immunocyte. Transcriptome data and methylation data from TCGA was used for correlation analysis, and GSE163720 from GEO was used for verification. Spearman's correlation test was performed to screen the gene with the highest correlation with infiltration level of the target immunocyte (Bishara and Hittner, 2012). And the correlation coefficient greater than 0.3 is considered vital (Funder and Ozer, 2019). Then, Kaplan-Meier survival analysis was used to explore the relationship between

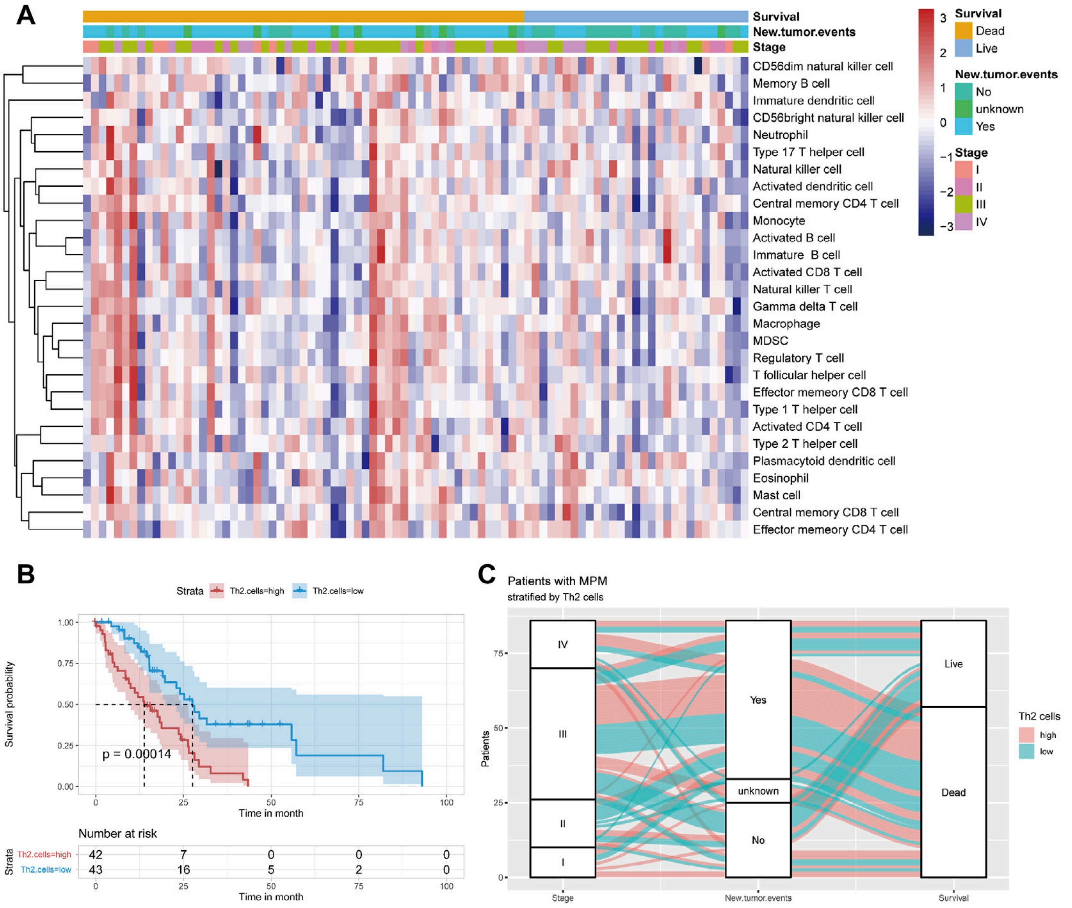


FIGURE 2 | (A) Immunocyte infiltration levels of 85 samples from TCGA. **(B)** Kaplan–Meier curve of patients from TCGA when using the median of Th2 cell infiltration levels as the cut-off value **(C)** Sankey plot depicting the relationship across the tumor stage, new tumor event and survival status of patients stratified by median value of Th2 cell infiltration levels.

transcriptome data and methylation data of the screened gene and prognosis of patients with MPM.

Immunohistochemistry Staining

MPM and normal pleural tissues were obtained from patients who undergone surgery in Sun Yat-Sen University Cancer Center. All specimens were diagnosed as MPM by pathologist in Sun Yat-Sen University Cancer Center.

Immunohistochemistry (IHC) staining was used to examine NFE2L3 expression in MPM tissues and paired normal pleural tissues. All paraffin-embedded specimens were cut into 5-μm sections and placed on glass slides, then baked at 60°C for 1 h. Firstly, dewaxing all specimens with xylene and rehydrating with ethanol, then immerse the specimens in sodium citrate-EDTA buffer, and using microwave heating for antigen retrieval. Secondly, using 3% hydrogen peroxide to inactivate endogenous peroxidase, and then blocking non-specific binding with 10% goat serum. Thirdly, incubating the slides with anti-NFE2L3 rabbit polyclonal antibody (1:200; NBP2-30870; Novus Biologicals) overnight at 4°C, after washing with PBS for 4 times, then adding secondary antibody polymer horseradish peroxidase to sections. Finally, the slices wer stained with DAB (3′-diaminobenzidine) and hematoxylin sequentially, dehydrated with gradient ethanol and mounted with neutral resin.

Pan-Cancer Analysis

TIMER (Tumor IMmune Estimation Resource) is a website that can provide a comprehensive analysis of transcriptome data from TCGA (Li et al., 2020). First, we study the differential expression of the screened gene between tumor and normal tissues across all TCGA samples using "DiffExp" module in TIMER. Then, using "Correlation" module, we performed correlation analysis of two related genes not only in MESO but also in many other cancer types. In this step, correlation analysis was adjusted by tumor purity.

RESULTS

Patients With High Th2 Cell Infiltration Suffer a Poor Prognosis

The heat map of the infiltration level of 28 immunocytes is shown in **Figure 2A**. Through Kaplan-Meier survival analysis, it can be

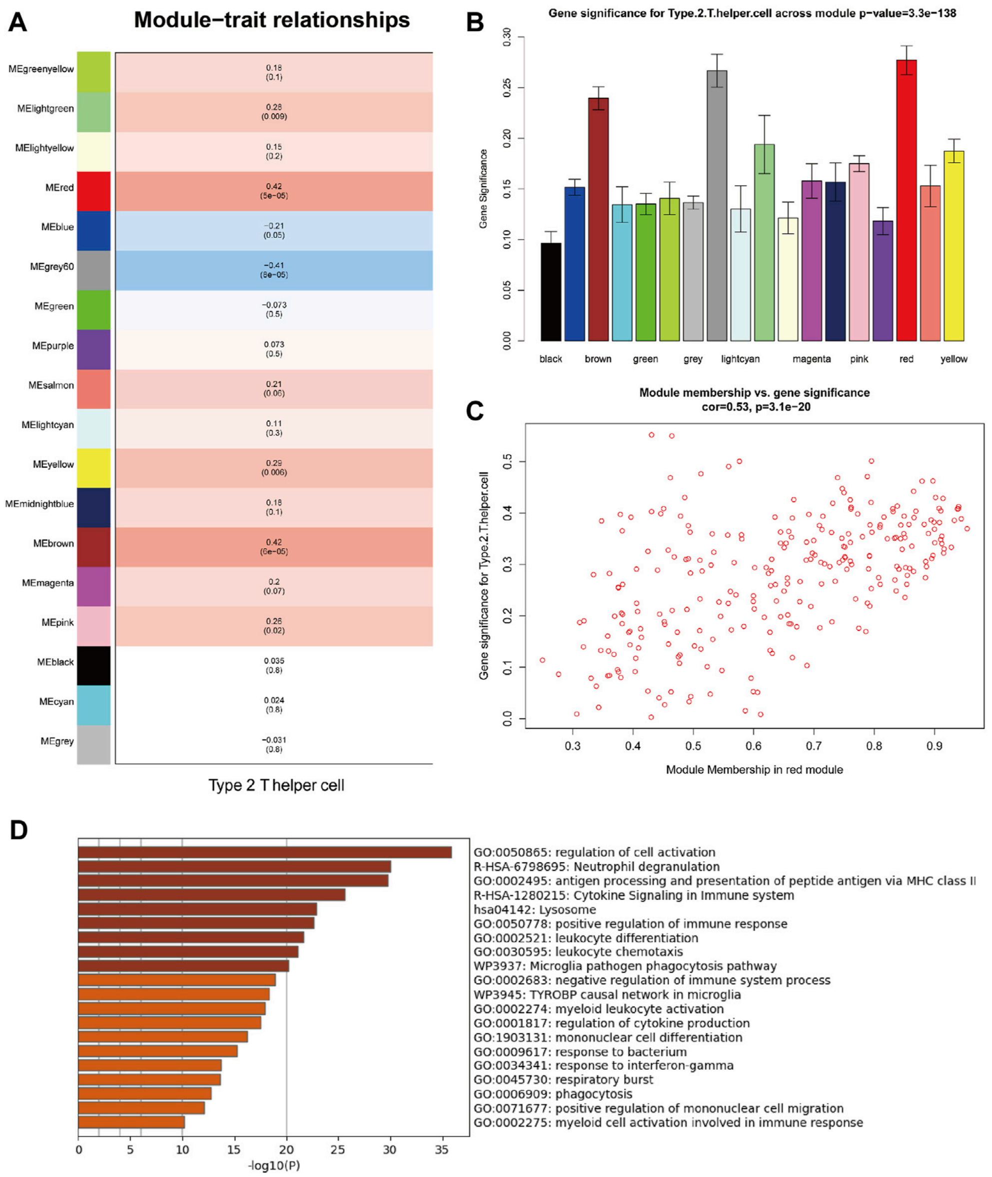


FIGURE 3 | Weighted gene co-expression network analysis (WGCNA) of MPM samples from TCGA. **(A)** Correlations between module eigengenes (MEs) and Th2 cell infiltration levels. **(B)** Gene significance (GS) for Th2 cell infiltration levels across all modules. **(C)** Gene significance (GS) of the genes contained in the red module versus module membership (Th2 cell infiltration levels). **(D)** Functional enrichment analysis of genes contained in the red module (Functional enrichment terms colored by *p*-values).

seen that patients with high Th2 cell infiltration suffer a significantly poor prognosis (**Figure 2B**, $p = 0.00014$). A Sankey plot presents the correlation between stage, recurrence and survival status of patients stratified by Th2 cell infiltration level (**Figure 2C**).

Identification of Co-Expressed Genes Associated With Th2 Cell

WGCNA can identify co-expressed genes that are highly correlated with Th2 cell infiltration. With the help of a dynamic tree-cutting algorithm, the 5,000 genes with the highest average expression were divided into 18 module eigengenes (MEs). Then Pearson's correlation coefficient was used to calculate the correlation between the MEs and clinicopathological traits. **Figures 3A,B** shows that the red module exhibits the highest correlation coefficient with Th2 cells (*Cor* = 0.42, *P* = 5e-05). The correlation across each gene in the red module and Th2 cells is plotted in **Figure 3C** (*Cor* = 0.53, *p* = 3.1e-20).

Functional enrichment analysis revealed that genes contained in the red module are mainly involved in various functions of the immune system (**Figure 3D**), including regulation of cell activation, neutrophil degranulation, antigen processing and

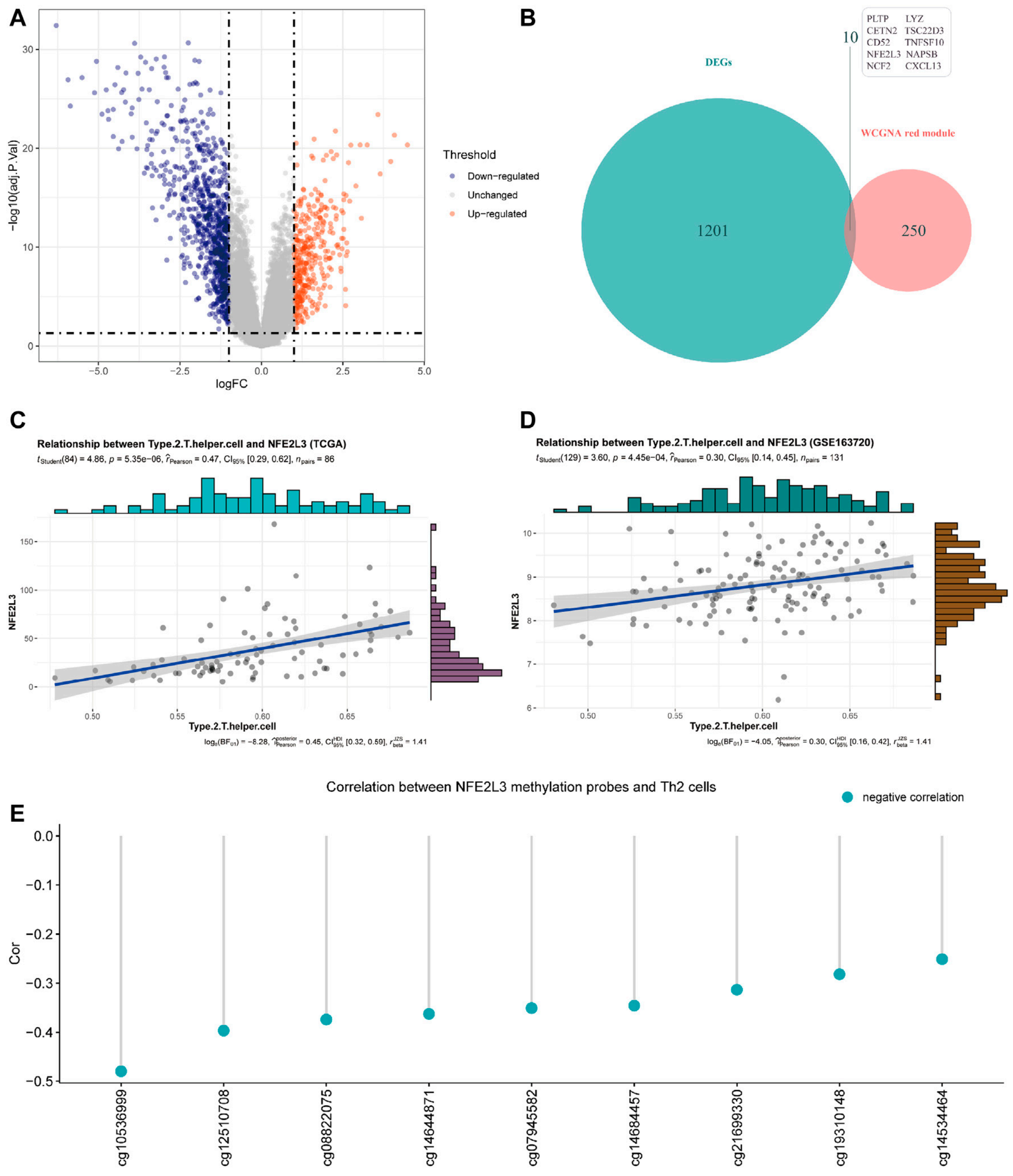


FIGURE 4 | (A) Volcano plot of differentially expressed genes (DEGs) between tumor tissues and paired normal tissues of samples from GSE51024 (adjusted p value < 0.05 and |logFC| >1). **(B)** Venn plot of the intersection of DEGs from GSE51024 and genes contained in red module from WGCNA **(C)** Correlation between NFE2L3 expression and Th2 cell infiltration levels of samples from TCGA. **(D)** Correlation between NFE2L3 expression and Th2 cell infiltration levels of samples from GSE163720 **(E)** Correlations between DNA methylation levels of NFE2L3 related sites and Th2 cell infiltration levels of samples from TCGA.

presentation of peptide antigen via MHC class II, cytokine signaling in immune system, lysosome, positive regulation of immune response, leukocyte differentiation, leukocyte chemotaxis, microglia pathogen phagocytosis pathway and negative regulation of immune system process.

Identification of Th2 Cell Infiltration-Related Genes

Differentially expressed genes (DEGs) analysis was performed to identify up-regulated and down-regulated genes between MPM tissues and normal lung parenchyma tissues. Through DEGs

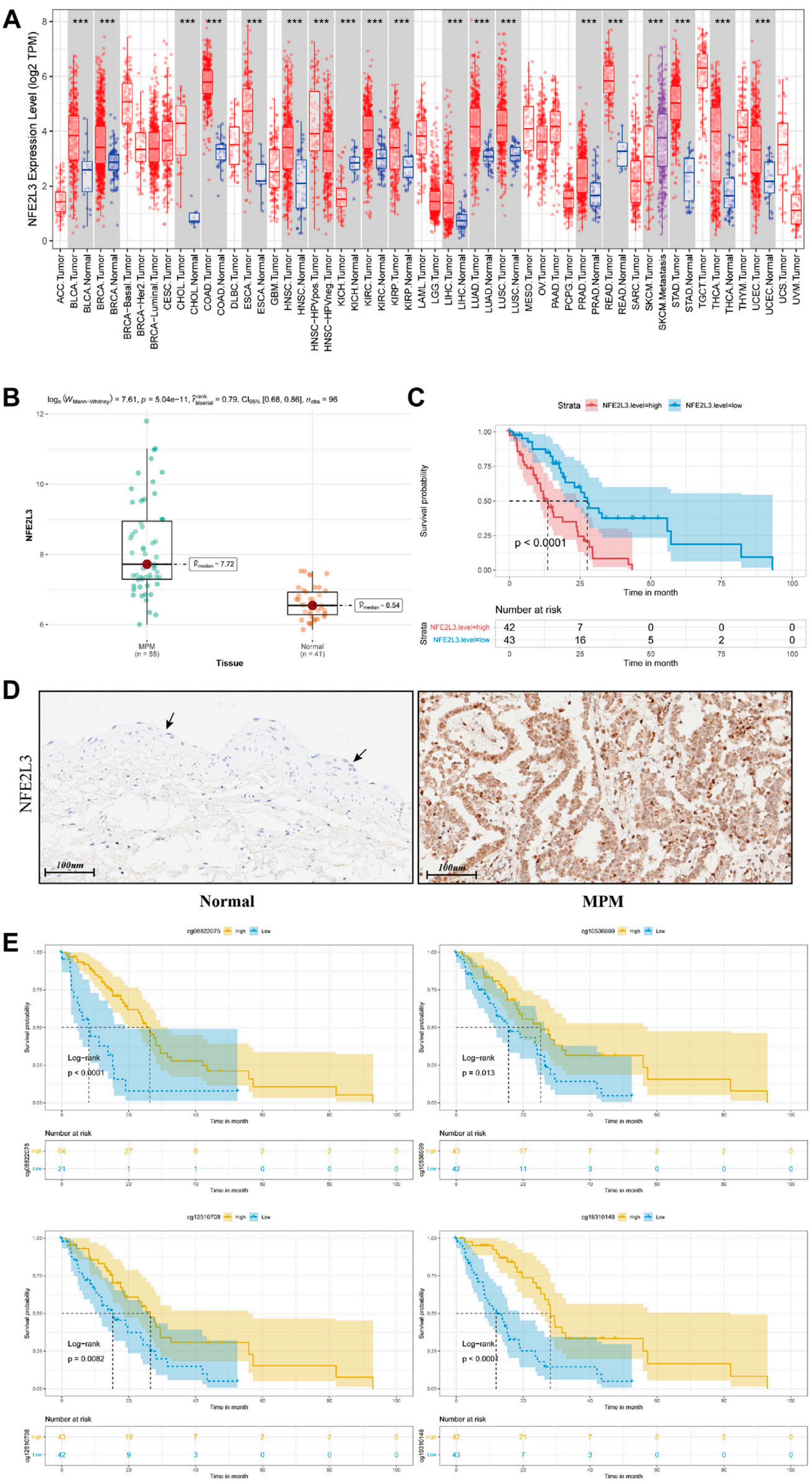


FIGURE 5 | (A) Pan-cancer analysis of NFE2L3 expression of all samples in TCGA. **(B)** Box plot of NFE2L3 expression in tumor tissues and paired normal tissues of samples from GSE51024 **(C)** Kaplan–Meier curve of patients from TCGA when using the median of NFE2L3 expression as the cut-off value. **(D)** Immunohistochemical staining of NFE2L3 in normal pleural epithelial tissue (black arrow) and malignant pleural mesothelioma tumor tissue **(E)** Kaplan–Meier curves depicting the relationships between DNA methylation levels of NFE2L3 related sites and prognosis of patients with MPM.

TABLE 1 | The basic information of DNA methylation sites of NFE2L3 and the correlations between their methylation levels and Th2 cell infiltration. Chrom, chromosome; ChromStart, starting position in the chromosome; ChromEnd, end position in the chromosome.

	Gene	Chrom	ChromStart	ChromEnd	Correlation Between NFE2L3 Methylation Probes and Th2 Cells	*p*-Value
					Cor	
cg03781084	NFE2L3	chr7	26,152,106	26,152,107	–0.110	3.15E-01
cg03886242	NFE2L3	chr7	26,152,412	26,152,413	–0.019	8.63E-01
cg04995722	NFE2L3	chr7	26,152,414	26,152,415	0.009	9.36E-01
cg07876897	NFE2L3	chr7	26,152,076	26,152,077	–0.103	3.45E-01
cg07945582	NFE2L3	chr7	26,166,959	26,166,960	–0.351	9.19E-04*
cg07986525	NFE2L3	chr7	26,152,579	26,152,580	–0.166	1.26E-01
cg08822075	NFE2L3	chr7	26,153,987	26,153,988	–0.374	3.86E-04*
cg10536999	NFE2L3	chr7	26,153,489	26,153,490	–0.480	2.96E-06*
cg12510708	NFE2L3	chr7	26,154,185	26,154,186	–0.397	1.55E-04*
cg13118545	NFE2L3	chr7	26,151,979	26,151,980	–0.110	3.15E-01
cg13855897	NFE2L3	chr7	26,186,769	26,186,770	–0.002	9.82E-01
cg14534464	NFE2L3	chr7	26,152,013	26,152,014	–0.251	1.97E-02*
cg14644871	NFE2L3	chr7	26,153,136	26,153,137	–0.363	5.96E-04*
cg14684457	NFE2L3	chr7	26,153,346	26,153,347	–0.346	1.11E-03*
cg16882373	NFE2L3	chr7	26,151,838	26,151,839	–0.123	2.58E-01
cg18844118	NFE2L3	chr7	26,151,869	26,151,870	–0.043	6.95E-01
cg19310,148	NFE2L3	chr7	26,156,654	26,156,655	–0.282	8.54E-03*
cg21699330	NFE2L3	chr7	26,153,412	26,153,413	–0.314	3.29E-03*
cg24424745	NFE2L3	chr7	26,160,252	26,160,253	–0.051	6.39E-01

analysis of the transcriptome data of GSE51024, 1,211 differentially expressed genes were screened, including 364 genes up-regulated in MPM tissues and 847 down-regulated genes. The expression profile of DEGs is shown in a volcano map (**Figure 4A**, adjusted p value <0.05 and $|\text{logFC}| >1$).

The red module obtained by WGCNA of data from TCGA contains 260 genes, of which 10 genes are differentially expressed in MPM tissues (**Figure 4B**). Through correlation analysis with Th2 cell infiltration level of samples from TCGA, NFE2L3 was identified that its expression level has the highest correlation with Th2 cell infiltration level (**Figure 4C**, *Cor* = 0.47). Then we utilized the transcriptome data of GSE163720 to validate and confirmed the strong correlation between NFE2L3 and Th2 cells (**Figure 4D**, *Cor* = 0.30).

By comparing MPM tissues and normal lung parenchyma tissues of GSE51024, NFE2L3 is highly expressed in MPM tissues (**Figure 5B**, p = 5.04e-11). And using tissue sections for immunohistochemical staining, we can observe that NFE2L3 is mainly expressed in tumor cell nuclei, but not in normal pleural tissues (**Figure 5D**). It is also confirmed by analyzing the transcriptome data of TCGA that patients with higher NFE2L3 expression suffer a worse prognosis (**Figure 5C**, $p <$ 0.0001).

In TCGA, DNA methylation levels of MPM tissues were determined with the Illumina Infinium Methylation 450 K array. After matching sample names, 9 of the 19 methylation sites of NFE2L3 were found to be correlated with the infiltration of Th2 cells, and all of them were negatively correlated (**Figure 4E**). **Table 1** details the basic information of NFE2L3 methylation sites and their correlations with Th2 cell infiltration level. Kaplan-Meier survival analysis also identified that the methylation levels of cg08822075, cg10536999, cg12510708, and cg19310148, were associated with the prognosis of patients, and patients with higher methylation levels had better prognosis (**Figure 5D**).

NFE2L3 Could Promote Th2 Cell Differentiation via IL-2/STAT5/NLRP3 Signaling Pathway

Through pan-cancer analysis of samples in TCGA, it can be seen that compared with normal tissues, NFE2L3 is highly expressed in almost all kinds of tumor tissues (**Figure 5A**).

According to previous research, NLRP3 is a transcriptional regulator of Th2 cell differentiation, and signal transducer IL-2R and STAT5 triggers its expression (Bruchard et al., 2015). Therefore, we further studied the correlation between NFE2L3 and the expression of related genes involved in this pathway.

As shown in **Table 2**, in MESO, NFE2L3 has a strong correlation with IL-2RA (*Cor* = 0.295), IL-2RB (*Cor* = 0.303) and IL-2RG (*Cor* = 0.228), which constitute the high-affinity IL2 receptor (IL-2R) (Wang et al., 2005). STAT5A and STAT5B are part of JAK/STAT signaling pathway, they can mediate transcriptional signals by forming homodimers or heterodimers (Maurer et al., 2019). In MESO, NFE2L3 and STAT5B show a strong correlation (**Table 2**, *Cor* = 0.35), but the correlation between NFE2L3 and STAT5A is not statistically significant. And there is also a strong correlation between NFE2L3 and NLRP3 (**Table 2**, *Cor* = 0.254). The potential interactions across NFE2L3, IL-2R, STAT5, NLRP3, Th2 cell and tumor cells are shown in **Figure 6**.

TABLE 2 | Correlation analyses between NFE2L3 and IL- 2R/STAT5/NLRP3 related genes in TIMER, which were adjusted by tumor purity. MESO, mesothelioma; BRCA (Her2), Her2 positive breast invasive carcinoma; DLBC, diffuse large B-cell lymphoma; ESCA, esophageal carcinoma; HNSC-HPVpos, HPV positive head and neck cancer; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LGG, low grade glioma; LIHC, liver hepatocellular carcinoma; LUSC, lung squamous cell carcinoma; PRAD, prostate adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; THCA, thyroid carcinoma; Cor, R value of Spearman's correlation.

		NFE2L3					
		IL2RA	**IL2RB**	**IL2RG**	**STAT5A**	**STAT5B**	**NLRP3**
MESO	Cor	0.295	0.303	0.228	0.085	0.35	0.254
	p-value	6.13E-03*	4.75E-03*	3.58E-03*	4.40E-01	1.03E-03*	1.89E-02*
BRCA (Her2)	Cor	0.443	0.491	0.459	0.49	0.315	0.417
	p-value	5.04E-04*	8.94E-05*	2.86E-04*	9.51E-05*	1.59E-02*	1.12E-03*
DLBC	Cor	0.418	0.719	0.311	0.098	0.587	0.638
	p-value	6.60E-03*	1.22E-07*	4.80E-02*	5.42E-01	5.57E-05*	7.29E-06*
ESCA	Cor	0.249	0.228	0.435	0.377	0.289	0.251
	p-value	7.63E-04*	2.10E-03*	1.06E-09*	1.89E-07*	8.41E-05*	6.64E-04*
HNSC (HPVpos)	Cor	0.456	0.612	0.671	0.566	0.227	0.275
	p-value	7.24E-06*	1.83E-10*	6.12E-13*	7.39E-09*	3.24E-02*	9.11E-03*
KIRC	Cor	0.317	0.423	0.355	0.359	-0.006	0.357
	p-value	3.15E-12*	1.74E-21*	4.05E-15*	1.73E-15*	9.03E-01	2.89E-15*
KIRP	Cor	0.33	0.344	0.304	0.396	0.434	0.391
	p-value	5.65E-08*	1.34E-08*	6.16E-07*	4.12E-11*	3.03E-13	7.41E-11*
LGG	Cor	0.169	0.294	0.324	0.427	0.241	0.28
	p-value	2.10E-04*	5.54E-11*	3.77E-13*	1.48E-22*	9.89E-08	4.40E-10*
LIHC	Cor	0.496	0.532	0.573	0.45	0.324	0.551
	p-value	7.62E-23*	1.49E-26*	1.61E-31*	1.32E-18*	7.51E-10*	9.91E-29*
LUSC	Cor	0.229	0.27	0.322	0.328	0.168	0.158
	p-value	4.38E-07*	2.02E-09*	5.81E-13*	1.99E-13*	2.31E-04*	5.36E-04*
PRAD	Cor	0.297	0.349	0.344	0.166	0.278	0.301
	p-value	6.04E-10*	2.30E-13*	5.43E-13*	7.02E-04*	7.89E-09*	3.90E-10*
SARC	Cor	0.249	0.395	0.401	0.207	0.062	0.432
	p-value	8.42E-25*	1.55E-10*	8.06E-11*	1.12E-03*	3.31E-01	1.64E-12*
SKCM	Cor	0.43	0.225	0.187	0.003	0.304	0.421
	p-value	5.84E-22*	1.16E-06*	5.56E-05*	9.55E-01	3.23E-11*	4.78E-21*
THCA	Cor	0.632	0.543	0.54	0.312	0.03	0.398
	p-value	9.24E-56*	8.45E-39*	2.83E-38*	1.86E-12*	5.12E-01	5.66E-20*

In addition to MESO, NFE2L3 also shows powerful correlations with IL-2RA, IL-2RB and IL-2RG in many other tumors, including BRCA-Her2, DLBC, ESCA, HNSC-HPVpos, KIRC, KIRP, LGG, LIHC, LUSC, PRAD, SARC, SKCM and THCA (**Table 2**). And among the above cancers, NFE2L3 shows a strong correlation with STAT5A or STAT5B as well (**Table 2**). Among the cancers analyzed, the correlations between NFE2L3 and NLRP3 are also statistically significant, among which the correlation is relatively low in LUSC (**Table 2**, *Cor* = 0.158), and the correlation is highest in DLBC (**Table 2**, *Cor* = 0.638).

DISCUSSION

As a very aggressive malignant tumor, malignant pleural mesothelioma (MPM) is believed to be closely related to asbestos exposure, BRCA1-associated protein 1 (BAP1) mutation and ionizing radiation to chest (Carbone et al., 2019). And a variety of somatic mutations including BAP1, TP53, NF2 and LATS1/2, are closely related to the occurrence and development of MPM (Bueno et al., 2016; Hmeljak et al., 2018; Yang et al., 2020a; Yang et al., 2020b). Due to the low incidence of MPM and the difficulty in diagnosis, there is still no unified and effective model for its treatment. Surgery used to be the only treatment for MPM, but the indications, extent of surgical resection and survival benefits of surgery are still controversial (Schipper et al., 2008; Cao et al., 2014). Beginning in 2004, the US Food and Drug Administration (FDA) approved the combination of cisplatin and pemetrexed as the first-line regimen for treatment of mesothelioma (Vogelzang et al., 2003). Although it has dramatically improved the survival of patients, the median survival period of patients who received combined chemotherapy after surgery still hovered between 17 and 25 months (Tsao et al., 2018).

Of note, with the advent of the age of immunotherapy, a variety of immune checkpoint inhibitors (ICIs) have brought a new dawn to the treatment of MPM. The DREAM study (Durvalumab with First-Line Chemotherapy in Mesothelioma) investigated the combination of PD-L1 inhibitor durvalumab and first-line chemotherapy (cisplatin and pemetrexed), and has brought survival benefits to patients (Nowak et al., 2020). Then a subsequent international, randomized, phase 3 study (CheckMate743) investigated Nivolumab in combination with Ipilimumab versus Pemetrexed with Cisplatin or Carboplatin as first line treatment in unresectable MPM, and identified that dual immune checkpoint inhibitors could bring long-term survival for these patients regardless of histological type (Baas et al., 2021).

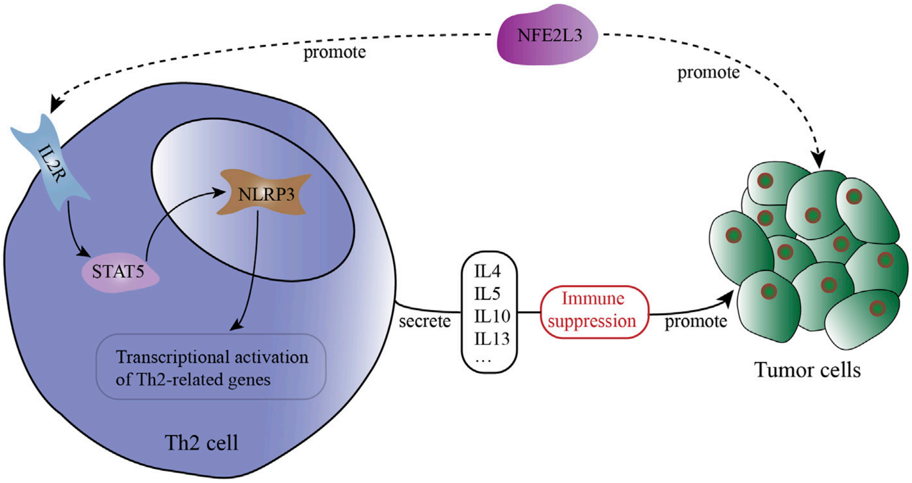


FIGURE 6 | The potential interactions across NFE2L3, IL-2R, STAT5, NLRP3, Th2 cell and tumor cells in the tumor microenvironment (TME).

Therefore, a better understanding of the tumor immune microenvironment and the construction of a more precise immune regulatory network will bring more individualized immunotherapy and survival benefits to patients. However, the tumor immune microenvironment is shaped by tumor cells and immunocytes together, and is in dynamic change (Schreiber et al., 2021). It is well known that $CD4^+$ T cell populations are abundant in this environment, including pro-tumor $CD4^+$ regulatory T cells (Tregs) and anti-tumor Th1 cells, but the role of Th2 cells is not clear so far (Facciabene et al., 2012; Tay et al., 2021).

In our study, the patients with high Th2 cell infiltration levels suffer a poor prognosis, which is consistent with some previous studies that Th2 cells are associated with tumor progression and poor prognosis in many cancers such as pancreatic cancer, breast cancer and melanoma (Nevala et al., 2009; De Monte et al., 2011; Zhang et al., 2015). The promotion effect of Th2 cells on tumors is probably due to the cytokines secreted by them, including IL-4, IL-5, IL-10 and IL-13 (Lee, 2014; Mollazadeh et al., 2019). IL-4 and IL-13 are highly similar in structure and function (Shi et al., 2021). After binding to their receptors, they can promote the proliferation, adhesion and metastasis of tumor cells, and may become potential targets for tumor treatment (Suzuki et al., 2015; Ghilardi et al., 2020). IL-5 can create a tumor-promoting immune microenvironment locally by recruiting eosinophils, thereby promoting the metastasis of tumor cells (Zaynagetdinov et al., 2015; Reichman et al., 2016). IL-10 can create an immunosuppressive tumor microenvironment through multiple pathways including NF-κB, and promote the transformation of cancer stemness (Yang et al., 2019; Saraiva et al., 2020).

Through further analysis, both the mRNA expression level and DNA methylation level of NFE2L3 was found to be highly correlated with the infiltration level of Th2 cells. NFE2L3 is a family member of the Cap'n'collar (CNC) transcription factors, and this family also include NFE2L1, NFE2L2, NF-E2, Bach1 and Bach2 (Ren et al., 2020). Among them, the family member that is widely investigated is NFE2L2, which has been confirmed as a driver gene of malignant tumor (DeNicola et al., 2011). As a homolog of NFE2L2, NFE2L3 has been proven to be related to multiple phenotypes of malignant tumors as well, including proliferation and epithelial-mesenchymal transition (EMT) (Bury et al., 2019; Ren et al., 2020). In our study, NFE2L3 was detected to be highly expressed in MPM tumor tissues, and the higher expression level is associated with poor prognosis. And the hypermethylation of multiple sites of NFE2L3 was also associated with better prognosis of patients with MPM. In multiple previous studies, somatic mutations of NFE2L2, the homolog of NFE2L3, was detected in plasma cell-free DNA (cfDNA) in hepatocellular carcinoma (HCC) and lung squamous cell carcinoma (LUSC), and was regarded as a non-invasive biomarker for tumor risk prediction and overall survival (Jeong et al., 2017; Jiao et al., 2021). Therefore, through next-generation sequencing (NGS) of tumor tissues and detection of cfDNA in peripheral blood, NFE2L3 may serve as a potential marker for the diagnosis and prognosis prediction of patients with MPM.

Then, we explored the underlying mechanism between NFE2L3 expression and Th2 cell differentiation. Th2 cell differentiation can be activated and modulated by a variety of regulators, including IL-4/STAT6 signaling pathway, IL-2/STAT5 signaling pathway and transcriptional regulator NLRP3, of which NLRP3 expression is triggered via IL-2/STAT5 signaling pathway (Lee, 2014; Bruchard et al., 2015; Jones et al., 2020). NLRP3 inflammasome is a kind of cytoplasmic protein complex, which has been proven that it can recruit myeloid-derived suppressor cells (MDSCs) and tumor-associated macrophages (TAMs) to promote tumor progression and metastasis (Moossavi et al., 2018; Hamarsheh and Zeiser, 2020). In the process of Th2 cells differentiation, NLRP3 was found to be localized in the nucleus and act as a transcription factor for Th2 cells (Bruchard et al., 2015).

In our study, the expression of NFE2L3 shows strong positive correlations with the expression of IL-2 receptor-related genes, STAT5 related genes and NLRP3. These correlations can be observed in multiple cancers, which are even more significant. Taken together, we speculated that NFE2L3, a novel biomarker in malignant pleural mesothelioma, can promote Th2 cell differentiation *via* IL-2/STAT5/NLRP3 signaling pathway in mesothelioma and many other cancers.

We acknowledge that there are several limitations in this study. In this article, our analysis is based only on transcriptome and DNA methylation data from TCGA and GEO without biological validation. Therefore, complete biological experiments are urgently needed to verify our conclusions in the future.

CONCLUSION

In our study, with the help of transcriptome data from multiple databases and a variety of bioinformatics analysis methods, we found that Th2 cell is a poor prognostic factor for patients with MPM. Through further screening, we found that NFE2L3 was highly expressed in tumor tissues of patients with MPM and both its mRNA expression level and DNA methylation level was highly correlated with the infiltration level of Th2 cells. Moreover, the correlation analysis in multiple cancers indicated that NFE2L3 was strongly correlated with the expression level of IL-2RA, IL-2RB, IL-2RG, STAT5A, STAT5B and NLRP3, which constitute the IL-2/STAT5/NLRP3 signaling pathway. Therefore, we hypothesize that NFE2L3, a novel biomarker in malignant pleural mesothelioma, may promote the differentiation of Th2 cells through the IL-2/STAT5/NLRP3 signaling pathway in multiple cancers.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article **Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

PL conceived this study; BL and PD retrieved and disposed data from GEO database; ZW and HY performed data analysis and prepared the manuscript; PL revised the manuscript. All authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank the contributors to the profile of GSE51024 and GSE163720 in the GEO datasets (https://www.ncbi.nlm.nih.gov/geoprofiles).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.805256/full#supplementary-material

REFERENCES

Baas, P., Scherpereel, A., Nowak, A. K., Fujimoto, N., Peters, S., Tsao, A. S., et al. (2021). First-line Nivolumab Plus Ipilimumab in Unresectable Malignant Pleural Mesothelioma (CheckMate 743): a Multicentre, Randomised, Open-Label, Phase 3 Trial. *The Lancet* 397 (10272), 375–386. doi:10.1016/S0140-6736(20)32714-8

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: Archive for Functional Genomics Data Sets-Update. *Nucleic Acids Res.* 41, D991–D995. doi:10.1093/nar/gks1193

Bishara, A. J., and Hittner, J. B. (2012). Testing the Significance of a Correlation with Nonnormal Data: Comparison of Pearson, Spearman, Transformation, and Resampling Approaches. *Psychol. Methods* 17 (3), 399–417. doi:10.1037/a0028087

Bruchard, M., Rebé, C., Derangère, V., Togbé, D., Ryffel, B., Boidot, R., et al. (2015). The Receptor NLRP3 Is a Transcriptional Regulator of TH2 Differentiation. *Nat. Immunol.* 16 (8), 859–870. doi:10.1038/ni.3202

Bueno, R., Stawiski, E. W., Goldstein, L. D., Durinck, S., De Rienzo, A., Modrusan, Z., et al. (2016). Comprehensive Genomic Analysis of Malignant Pleural Mesothelioma Identifies Recurrent Mutations, Gene Fusions and Splicing Alterations. *Nat. Genet.* 48 (4), 407–416. doi:10.1038/ng.3520

Bury, M., Le Calvé, B., Lessard, F., Dal Maso, T., Saliba, J., Michiels, C., et al. (2019). NFE2L3 Controls Colon Cancer Cell Growth through Regulation of DUX4, a CDK1 Inhibitor. *Cel Rep.* 29 (6), 1469–1481. doi:10.1016/j.celrep.2019.09.087

Cancer Genome Atlas ResearchWeinstein, J. N., Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., et al. (2013). The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat. Genet.* 45 (10), 1113–1120. doi:10.1038/ng.2764

Cao, C., Tian, D., Park, J., Allan, J., Pataky, K. A., and Yan, T. D. (2014). A Systematic Review and Meta-Analysis of Surgical Treatments for Malignant Pleural Mesothelioma. *Lung Cancer* 83 (2), 240–245. doi:10.1016/j.lungcan.2013.11.026

Carbone, M., Adusumilli, P. S., Alexander, H. R., Jr., Baas, P., Bardelli, F., Bononi, A., et al. (2019). Mesothelioma: Scientific Clues for Prevention, Diagnosis, and Therapy. *CA A. Cancer J. Clin.* 69 (5), 402–429. doi:10.3322/caac.21572

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cel Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

de Gooijer, C. J., Baas, P., and Burgers, J. A. (2018). Current Chemotherapy Strategies in Malignant Pleural Mesothelioma. *Transl. Lung Cancer Res.* 7 (5), 574–583. doi:10.21037/tlcr.2018.04.10

De Monte, L., Reni, M., Tassi, E., Clavenna, D., Papa, I., Recalde, H., et al. (2011). Intratumor T Helper Type 2 Cell Infiltrate Correlates with Cancer-Associated Fibroblast Thymic Stromal Lymphopoietin Production and Reduced Survival in Pancreatic Cancer. *J. Exp. Med.* 208 (3), 469–478. doi:10.1084/jem.20101876

DeNicola, G. M., Karreth, F. A., Humpton, T. J., Gopinathan, A., Wei, C., Frese, K., et al. (2011). Oncogene-induced Nrf2 Transcription Promotes ROS

Detoxification and Tumorigenesis. *Nature* 475 (7354), 106–109. doi:10.1038/nature10189

Facciabene, A., Motz, G. T., and Coukos, G. (2012). T-regulatory Cells: Key Players in Tumor Immune Escape and Angiogenesis: Figure 1. *Cancer Res.* 72 (9), 2162–2171. doi:10.1158/0008-5472.CAN-11-3687

Fernandez-Cuesta, L., Mangiante, L., Alcala, N., and Foll, M. (2021). Challenges in Lung and Thoracic Pathology: Molecular Advances in the Classification of Pleural Mesotheliomas. *Virchows Arch.* 478 (1), 73–80. doi:10.1007/s00428-020-02980-9

Funder, D. C., and Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Adv. Methods Practices Psychol. Sci.* 2, 156–168. doi:10.1177/2515245919847202

Ghilardi, N., Pappu, R., Arron, J. R., and Chan, A. C. (2020). 30 Years of Biotherapeutics Development-What Have We Learned? *Annu. Rev. Immunol.* 38, 249–287. doi:10.1146/annurev-immunol-101619-031510

Hamarsheh, S. A., and Zeiser, R. (2020). NLRP3 Inflammasome Activation in Cancer: A Double-Edged Sword. *Front. Immunol.* 11, 1444. doi:10.3389/fimmu.2020.01444

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7

Hmeljak, J., Sanchez-Vega, F., Hoadley, K. A., Shih, J., Stewart, C., Heiman, D., et al. (2018). Integrative Molecular Characterization of Malignant Pleural Mesothelioma. *Cancer Discov.* 8 (12), 1548–1565. doi:10.1158/2159-8290.CD-18-0804

Jeong, Y., Hoang, N. T., Lovejoy, A., Stehr, H., Newman, A. M., Gentles, A. J., et al. (2017). Role of KEAP1/NRF2 and TP53 Mutations in Lung Squamous Cell Carcinoma Development and Radiation Resistance. *Cancer Discov.* 7 (1), 86–101. doi:10.1158/2159-8290.CD-16-0127

Jiao, J., Sanchez, J. I., Thompson, E. J., Mao, X., McCormick, J. B., Fisher-Hoch, S. P., et al. (2021). Somatic Mutations in Circulating Cell-free DNA and Risk for Hepatocellular Carcinoma in Hispanics. *Ijms* 22 (14), 7411. doi:10.3390/ijms22147411

Jones, D. M., Read, K. A., and Oestreich, K. J. (2020). Dynamic Roles for IL-2-STAT5 Signaling in Effector and Regulatory CD4+ T Cell Populations. *J.I.* 205 (7), 1721–1730. doi:10.4049/jimmunol.2000612

Kindler, H. L., Ismaila, N., Armato, S. G., 3rd, Bueno, R., Hesdorffer, M., Jahan, T., et al. (2018). Treatment of Malignant Pleural Mesothelioma: American Society of Clinical Oncology Clinical Practice Guideline. *Jco* 36 (13), 1343–1373. doi:10.1200/JCO.2017.76.6394

Kusuda, T., Shigemasa, K., Arihiro, K., Fujii, T., Nagai, N., and Ohama, K. (2005). Relative Expression Levels of Th1 and Th2 Cytokine mRNA Are Independent Prognostic Factors in Patients with Ovarian Cancer. *Oncol. Rep.* 13 (6), 1153–1158. doi:10.3892/or.13.6.1153

Langfelder, P., and Horvath, S. (2008). WGCNA: an R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* 9, 559. doi:10.1186/1471-2105-9-559

Lee, G. R. (2014). Transcriptional Regulation of T Helper Type 2 Differentiation. *Immunology* 141 (4), 498–505. doi:10.1111/imm.12216

Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., et al. (2020). TIMER2.0 for Analysis of Tumor-Infiltrating Immune Cells. *Nucleic Acids Res.* 48 (W1), W509–W514. doi:10.1093/nar/gkaa407

Lievense, L. A., Sterman, D. H., Cornelissen, R., and Aerts, J. G. (2017). Checkpoint Blockade in Lung Cancer and Mesothelioma. *Am. J. Respir. Crit. Care Med.* 196 (3), 274–282. doi:10.1164/rccm.201608-1755CI

Maurer, B., Kollmann, S., Pickem, J., Hoelbl-Kovacic, A., and Sexl, V. (2019). STAT5A and STAT5B-Twins with Different Personalities in Hematopoiesis and Leukemia. *Cancers* 11 (11), 1726. doi:10.3390/cancers11111726

Mollazadeh, H., Cicero, A. F. G., Blesso, C. N., Pirro, M., Majeed, M., and Sahebkar, A. (2019). Immune Modulation by Curcumin: The Role of Interleukin-10. *Crit. Rev. Food Sci. Nutr.* 59 (1), 89–101. doi:10.1080/10408398.2017.1358139

Moossavi, M., Parsamanesh, N., Bahrami, A., Atkin, S. L., and Sahebkar, A. (2018). Role of the NLRP3 Inflammasome in Cancer. *Mol. Cancer* 17 (1), 158. doi:10.1186/s12943-018-0900-3

Nevala, W. K., Vachon, C. M., Leontovich, A. A., Scott, C. G., Thompson, M. A., Markovic, S. N., et al. (2009). Evidence of Systemic Th2-Driven Chronic Inflammation in Patients with Metastatic Melanoma. *Clin. Cancer Res.* 15 (6), 1931–1939. doi:10.1158/1078-0432.CCR-08-1980

Nowak, A. K., Lesterhuis, W. J., Kok, P.-S., Brown, C., Hughes, B. G., Karikios, D. J., et al. (2020). Durvalumab with First-Line Chemotherapy in Previously Untreated Malignant Pleural Mesothelioma (DREAM): a Multicentre, Single-Arm, Phase 2 Trial with a Safety Run-In. *Lancet Oncol.* 21 (9), 1213–1223. doi:10.1016/S1470-2045(20)30462-9

Perme, M. P., and Manevski, D. (2019). Confidence Intervals for the Mann-Whitney Test. *Stat. Methods Med. Res.* 28 (12), 3755–3768. doi:10.1177/0962280218814556

Reichman, H., Karo-Atar, D., and Munitz, A. (2016). Emerging Roles for Eosinophils in the Tumor Microenvironment. *Trends Cancer* 2 (11), 664–675. doi:10.1016/j.trecan.2016.10.002

Ren, Y., Wang, Y., Hao, S., Yang, Y., Xiong, W., Qiu, L., et al. (2020). NFE2L3 Promotes Malignant Behavior and EMT of Human Hepatocellular Carcinoma (HepG2) Cells via Wnt/β-catenin P-athway. *J. Cancer* 11 (23), 6939–6949. doi:10.7150/jca.48100

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Saraiva, M., Vieira, P., and O'Garra, A. (2020). Biology and Therapeutic Potential of Interleukin-10. *J. Exp. Med.* 217 (1), e20190418. doi:10.1084/jem.20190418

Schipper, P. H., Nichols, F. C., Thomse, K. M., Deschamps, C., Cassivi, S. D., Allen, M. S., et al. (2008). Malignant Pleural Mesothelioma: Surgical Management in 285 Patients. *Ann. Thorac. Surg.* 85 (1), 257–264. doi:10.1016/j.athoracsur.2007.06.066

Schreiber, S., Hammers, C. M., Kaasch, A. J., Schraven, B., Dudeck, A., and Kahlfuss, S. (2021). Metabolic Interdependency of Th2 Cell-Mediated Type 2 Immunity and the Tumor Microenvironment. *Front. Immunol.* 12, 632581. doi:10.3389/fimmu.2021.632581

Shi, J., Song, X., Traub, B., Luxenhofer, M., and Kornmann, M. (2021). Involvement of IL-4, IL-13 and Their Receptors in Pancreatic Cancer. *Ijms* 22 (6), 2998. doi:10.3390/ijms22062998

Suzuki, A., Leland, P., Joshi, B. H., and Puri, R. K. (2015). Targeting of IL-4 and IL-13 Receptors for Cancer Therapy. *Cytokine* 75 (1), 79–88. doi:10.1016/j.cyto.2015.05.026

Tay, R. E., Richardson, E. K., and Toh, H. C. (2021). Revisiting the Role of CD4+ T Cells in Cancer Immunotherapy-New Insights into Old Paradigms. *Cancer Gene Ther.* 28 (1-2), 5–17. doi:10.1038/s41417-020-0183-x

Tsao, A. S., Lindwasser, O. W., Adjei, A. A., Adusumilli, P. S., Beyers, M. L., Blumenthal, G. M., et al. (2018). Current and Future Management of Malignant Mesothelioma: A Consensus Report from the National Cancer Institute Thoracic Malignancy Steering Committee, International Association for the Study of Lung Cancer, and Mesothelioma Applied Research Foundation. *J. Thorac. Oncol.* 13 (11), 1655–1667. doi:10.1016/j.jtho.2018.08.2036

Vogelzang, N. J., Rusthoven, J. J., Symanowski, J., Denham, C., Kaukel, E., Ruffie, P., et al. (2003). Phase III Study of Pemetrexed in Combination with Cisplatin versus Cisplatin Alone in Patients with Malignant Pleural Mesothelioma. *Jco* 21 (14), 2636–2644. doi:10.1200/JCO.2003.11.136

Wang, X., Rickert, M., and Garcia, K. C. (2005). Structure of the Quaternary Complex of Interleukin-2 with its α, SS, and γ C Receptors. *Science* 310 (5751), 1159–1163. doi:10.1126/science.1117893

Yang, H., Xu, D., Schmid, R. A., and Peng, R.-W. (2020a). Biomarker-guided Targeted and Immunotherapies in Malignant Pleural Mesothelioma. *Ther. Adv. Med. Oncol.* 12, 175883592097142. doi:10.1177/1758835920971421

Yang, H., Xu, D., Yang, Z., Yao, F., Zhao, H., Schmid, R. A., et al. (2020b). Systematic Analysis of Aberrant Biochemical Networks and Potential Drug Vulnerabilities Induced by Tumor Suppressor Loss in Malignant Pleural Mesothelioma. *Cancers* 12 (8), 2310. doi:10.3390/cancers12082310

Yang, L., Dong, Y., Li, Y., Wang, D., Liu, S., Wang, D., et al. (2019). IL-10 Derived from M2 Macrophage Promotes Cancer Stemness via JAK1/STAT1/NF-κB/Notch1 Pathway in Non-small Cell Lung Cancer. *Int. J. Cancer* 145 (4), 1099–1110. doi:10.1002/ijc.32151

Zaynagetdinov, R., Sherrill, T. P., Gleaves, L. A., McLoed, A. G., Saxon, J. A., Habermann, A. C., et al. (2015). Interleukin-5 Facilitates Lung Metastasis by Modulating the Immune Microenvironment. *Cancer Res.* 75 (8), 1624–1634. doi:10.1158/0008-5472.CAN-14-2379

Zhang, Q., Qin, J., Zhong, L., Gong, L., Zhang, B., Zhang, Y., et al. (2015). CCL5-Mediated Th2 Immune Polarization Promotes Metastasis in Luminal Breast Cancer. *Cancer Res.* 75 (20), 4312–4321. doi:10.1158/0008-5472.CAN-14-3590

Zhang, Z., Hernandez, K., Savage, J., Li, S., Miller, D., Agrawal, S., et al. (2021). Uniform Genomic Data Analysis in the NCI Genomic Data Commons. *Nat. Commun.* 12 (1), 1226. doi:10.1038/s41467-021-21254-9

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

Zhu, J., Yamane, W. E., and Paul, W. E. (2010). Differentiation of Effector CD4 T Cell Populations. *Annu. Rev. Immunol.* 28, 445–489. doi:10.1146/annurev-immunol-030409-101212

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.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 27 May 2022
doi: 10.3389/fgene.2022.820065

The Chemokines Initiating and Maintaining Immune Hot Phenotype Are Prognostic in ICB of HNSCC

Yuhong Huang[1†], Han Liu[1,2†], Xuena Liu[3†], Nan Li[1,2], Han Bai[1], Chenyang Guo[1], Tian Xu[1], Lei Zhu[1,2], Chao Liu[1,2] and Jing Xiao[1,2]**

[1]Department of Oral Pathology, School of Stomatology, Dalian Medical University, Dalian, China, [2]Dalian Key Laboratory of Basic Research in Oral Medicine, School of Stomatology, Dalian Medical University, Dalian, China, [3]Department of Nuclear Medicine, The 2nd Hospital Affiliated to Dalian Medical University, Dalian, China

Background: The immune checkpoint blockade (ICB) with anti-programmed cell death protein 1(PD-1) on HNSCC is not as effective as on other tumors. In this study, we try to find out the key factors in the heterogeneous tumor-associated monocyte/macrophage (TAMM) that could regulate immune responses and predict the validity of ICB on HNSCC.

Experimental Design: To explore the correlation of the TAMM heterogeneity with the immune properties and prognosis of HNSCC, we established the differentiation trajectory of TAMM by analyzing the single-cell RNA-seq data of HNSCC, by which the HNSCC patients were divided into different sub-populations. Then, we exploited the topology of the network to screen out the genes critical for immune hot phenotype of HNSCC, as well as their roles in TAMM differentiation, tumor immune cycle, and progression. Finally, these key genes were used to construct a neural net model via deep-learning framework to predict the validity of treatment with anti-PD-1/PDL-1

Results: According to the differentiation trajectory, the genes involved in TAMM differentiation were categorized into early and later groups. Then, the early group genes divided the HNSCC patients into sub-populations with more detailed immune properties. Through network topology, CXCL9, 10, 11, and CLL5 related to TAMM differentiation in the TME were identified as the key genes initiating and maintaining the immune hot phenotype in HNSCC by remarkably strengthening immune responses and infiltration. Genome wide, CASP8 mutations were found to be key to triggering immune responses in the immune hot phenotype. On the other hand, in the immune cold phenotype, the evident changes in CNV resulted in immune evasion by disrupting immune balance. Finally, based on the framework of CXCL9-11, CLL5, $CD8^+$, $CD4^+$ T cells, and Macrophage M1, the neural network model could predict the validity of PD-1/PDL-1 therapy with 75% of AUC in the test cohort.

Conclusion: We concluded that the CXCL9, 10,11, and CCL5 mediated TAMM differentiation and constructed immune hot phenotype of HNSCC. Since they positively regulated immune cells and immune cycle in HNSCC, the CXCL9-11 and CCL5 could be used to predict the effects of anti-PD-1/PDL-1 therapy on HNSCC.

Keywords: squamous cell carcinoma of head and neck (HNSCC), tumor-associated monocyte/macrophage (TAMM), immune checkpoint blockade (ICB), tumor micro-environment (TME), CXCL, CCL5, PD-1/PD-L1

Edited by:
Eszter Lakatos,
Queen Mary University of London,
United Kingdom

Reviewed by:
Daniel Roden,
Garvan Institute of Medical Research,
Australia
Ravindra Gujar,
Oncosec Immunotherapies,
United States

***Correspondence:**
Chao Liu
cliu@dmu.edu.cn
Jing Xiao
xiaoj@dmu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to
Cancer Genetics and Oncogenomics,
a section of the journal
Frontiers in Genetics

Received: 22 November 2021
Accepted: 28 March 2022
Published: 27 May 2022

Citation:
Huang Y, Liu H, Liu X, Li N, Bai H,
Guo C, Xu T, Zhu L, Liu C and Xiao J
(2022) The Chemokines Initiating and
Maintaining Immune Hot Phenotype
Are Prognostic in ICB of HNSCC.
Front. Genet. 13:820065.
doi: 10.3389/fgene.2022.820065

INTRODUCTION

Application of immune checkpoint blockage (ICB) has significantly improved the prognosis of multiple tumors, but exerted limited effects on head and neck squamous cell carcinoma (HNSCC) because less than 30% of the patients got a better prognosis (Curran et al., 2010; Topalian et al., 2012; Seiwert et al., 2016; Clarke et al., 2021; García Campelo et al., 2021). To find out the HNSCC sub-populations susceptible to ICB therapy, various criteria have been proposed for HNSCC classification, by which HNSCC was classified into the enhanced and decreased immune subtypes with the immune-related genes (Cao et al., 2018), into the $CD8^+$ high and $CD8^+$ low subtypes with the density of infiltrating $CD8^+$ T cells (Saloura et al., 2019), into the basal, mesenchymal, atypical, and classical subtypes with the integrated genomic characteristics (Walter et al., 2013), and even into the $m6A^{high}$ and $m6A^{low}$ subtypes with the N6-methyladenosine (m6A) methylation levels on mRNAs (Yi et al., 2020). Although these criteria explicated the clinical and immune characteristics of HNSCC from different perspectives, how these HNSCC subtypes formed and the roles of the genes involved in it remained elusive.

When inflammation took place, the tumor-associated monocyte/macrophages in circulating blood were motivated into the inflammatory focus to maintain homeostasis, eliminate pathogens, and balance immune responses (Shi & Pamer, 2011). There were three types of tumor associated monocyte/macrophages, namely, the classical ($CD14^+$; $CD16^-$), the non-classical ($CD16^+$), and the intermediate tumor associated monocyte/macrophages ($CD14^+$; $CD16^+$) (Ziegler-Heitbrock et al., 2010). During tumorigenesis, the tumor-associated monocyte/macrophages in circulation (mainly the classical type) were chemoattracted into tumor focus, and differentiated gradually into dendritic cells (DC) and Tumor Associated Macrophages (TAM) (Movahedi et al., 2010; Franklin et al., 2014; Guilliams et al., 2014; Li B et al., 2020). More than 50% of the immune cells in tumor micro-environment (TME) were TAM that affected the migration, invasion, angiogenesis, and drug-resistance of tumors (Watters et al., 2005; Kimura et al., 2007; Zheng et al., 2018). The M1-like phenotype of TAM exhibited the inhibitory effects on tumors, such as the promoted inflammatory response and chemoattraction of immune cells (Goswami et al., 2017). Reversely, M2-like phenotype of TAM suppressed inflammatory response, and enhanced immune evasion, angiogenesis, and metastasis, which resulted in a poor prognosis (Hu et al., 2016; Seminerio et al., 2018). However, recent studies reported that M1-like phenotype of TAM was also related to the poor prognosis in HNSCC and medulloblastoma by suppressing inflammatory response and promoting metastasis (Lee et al., 2018; Xiao et al., 2018). Previously, TAM was thought mainly to be macrophage M2, while the increasing evidence indicated that TAM also exhibited the phenotype of macrophage M1, suggesting that TAM contained the third population of macrophages other than macrophage M1 and M2. The third macrophage population was supposed to co-express the M1 and M2 characteristics and transform into M1 or M2 in certain instances (Estko et al., 2015; L. ; Gao et al., 2016; Kloepper et al., 2016). All the above findings indicated that the role of TAM in tumor progression could be complicated and not simply attributed to macrophage M1 and M2. Since the TAM was differentiated from the tumor-associated monocyte/macrophages gradually, the differentiating and differentiated TAM were termed as tumor-associated monocyte/macrophages/Macrophages (TAMM) in recent studies (Cassetta et al., 2019; Singhal et al., 2019). More and more studies implicated the TAMM as the potential target of ICB therapy. The relevance between TAMM responses and ICB therapy has been established by bioinformatic methods. In triple negative breast cancer, machine learning identified the TAMM-expressed genes which were highly associated with the prognosis and ICB therapy, and constructed a model predicting the response to ICB therapy with the 100% validation queue AUC (Bao et al., 2021). Moreover, WGCNA was used to find that the marker genes expressed in TAMM of glioblastoma, which were highly correlated with prognosis and ICB therapy, and were more active in the patients susceptible to ICB therapy (Zhang et al., 2021). Despite this, there are relatively few studies on HNSCC concerning the role of TAMM in immune response and ICB therapy. Since TAMM differentiation endowed TAMM with heterogeneity dynamically, instead of statically, we proposed a criterion that combined the genes involved in TAMM differentiation with the immune cells to depict the immune phenotype and prognosis of HNSCC in more detail.

MATERIALS AND METHODS

Data Collection

The single cell RNA-seq (scRNA-seq) data of GSE139324 (10X genomics), including the tumor infiltrating immune cells from 16 HPV negative patients and the immune cells from the peripheral blood of a healthy donor, and GSE103322 (Smart-seq2), containing 5,902 single cells from 18 HNSCC patients, were obtained from Gene Expression Omnibus (GEO). Multiomics data and clinical data of 502 HNSCC patients obtained from The Cancer Genome Atlas (TCGA) database (**Supplementary Table S1**), including mRNA expression (level 3, Illumina RNA-Seq), miRNA expression (level 3, Illumina miRNA-Seq), somatic copy number variation (CNV level 3, Affymetrix SNP 6.0), and somatic mutation (level 4, MAF files), were obtained from UCSC Xena browser. The array data and clinical data of five HNSCC cohorts, GSE65858 (n = 270), GSE40774 (n = 134), GSE39366 (n = 138), GSE117973 (n = 77), and GSE41613 (N = 97), were obtained from Gene Expression Omnibus (GEO) database (**Supplementary Table S1**). The bulk transcriptome data and clinical data of six cohorts accepted the PDL-1/PD-L1 antibody immunotherapy, namely GSE93157 (n = 65, Non-Small Cell Lung Carcinoma, HNSCC and Melanoma), GSE154538 (n = 8, gastrointestinal cancer), GSE141119 (n = 12, melanoma), GSE91061 (n = 109, melanoma and non-small cell lung cancer), GSE78220 (n = 28, melanomas), GSE176307 (n = 88, Metastatic Urothelial Cancer), and the IMvigor210 (n = 348, bladder cancer), were obtained from Gene Expression Omnibus (GEO) and the IMvigor210 database (**Supplementary Table S2**). GSE93157 was array data, while GSE154538, GSE141119, GSE91061, GSE78220, GSE176307, and the IMvigor210 were bulk transcriptome data. The mRNA-seq data from the HNSCC cell line that accepted the treatment of anti-tumor drugs were obtained from Genomics of Drug Sensitivity in Cancer (GDSC).

Data Processing

In the scRNA-seq from GSE139324, with the exclusion of the genes detected in fewer than three cells, the cells containing mRNA more than 4,500 or less than 200, and the cells expressing mitochondria genes more than 10% transcripts, there were 19,718 genes from 39,711 qualified cells of total 39,994 cells (283 cells were screened out). Through SCT in seurant package SCT, the data from the 18 patients in GSE139324 cohort were integrated to screen out the non-biological inferences, such as batch effect. Similarly, in the scRNA-seq of GSE103322 (Smart2-seq without screen), there were 21,519 genes and 5,844 qualified cells from 18 patients integrated by SCT. The HNSCC RNA-seq counts [$\log_2$(rawcounts+1)] obtained from USCS through exp [$\log_2$(rawcounts+1)-1] were restored to raw counts, and then the log2(fpkm-uq+1) from USCS was used to compare them with the data from other databases. There were 501 HNSCC samples (one normal sample was excluded) for the subsequent analyses. The GSE65858, GSE40774, GSE39366, GSE117973, and GSE41613 were normalized prior to following analyses. There were 501 samples in GSE93157, GSE154538, GSE141119, GSE91061, GSE78220, GSE176307, and the IMvigor210 for the following analyses except defective and reiterated data. The data from the RNA-seq of GDSC2 cell line were normalized with TPM for subsequent processing.

Analyses on Squamous Cell Carcinoma of Head and Neck scRNA-Seq Data

For the GSE139324 cohort: 1) the data integrated with "SCT" Seurat package was applied for PCA analysis to find out the first 50 principal component analysis (PCA); 2) Umap (Uniform Manifold Approximation and Projection for Dimension Reduction) dimension reduction was performed on the 50 PCAs. In this unsupervised clustering, the function of FindNeighbors in Seurat package was used to construct a KNN graph based on the Euclidean distance in PCA space (top 50 PCAs, k = 20), and then, the function of FindClusters (Louvain algorithm) was used to cluster the cells with the resolution of 0.1. The K-NN clustering classified the consequences undergoing the dimension reduction into four clusters, which were annotated by SingleR as NK cells (n = 14,925), T cells (n = 14,073), B cells (n = 2,789), and tumor-associated monocyte/macrophages cells (n = 7,924). 3) Tumor-associated monocyte/macrophages cells were classified by K-NN into seven further clusters. Cluster 0, 1, 2, and 4 were annotated by SingleR as tumor-associated monocyte/macrophages (n = 6,875), while the cluster3 (n = 312), 5 (n = 478), and 6 (n = 259) as T and B cells. 4) The T cells were applied for Multimodal reference mapping (Hao et al., 2021) and divided into eight clusters, namely, the CD4 CTL, CD4 Navie cells, CD4 TCM, CD4 TEM, CD8 Navie cells, CD8 TCM, CD8 TEM, and proliferating T cells (**Supplementary Table S3**). 5) The tumor-associated monocyte/macrophages were applied for GSVA analysis for function enrichment. 6) The tumor-associated monocyte/macrophages were applied for pseudotime analysis through Monocle2 package and Destiny package, which adopted different manners to reduce the dimensions of the high-dimensional data. The single cell was separated and projected into low-dimensional space to form a differentiation trajectory with knots. Each knot represented a similar status of differentiation. (1) Through the data of single cell lineage, Monocle 2 adopted the embedding converse diagraph to learn the explicit principal graph (Packer et al., 2019). 2) Destiny adopted the diffuse maps (differentiating cells follow noisy diffusion-like dynamics) to mimic the division from multipotent cells (Coifman et al., 2005). 7) The "InferCNV" R package 1.10.1 (Patel et al., 2014) and CellPhoneDB (Python edition) (Efremova et al., 2020) were performed on all clusters for CNV analysis (normal blood cells as control) and cell communication analysis. $\mathrm{CNV_k}(\mathrm{i}) = \sum_{\mathrm{j=i-50}}^{\mathrm{i+50}} \mathrm{E_k}(\mathrm{O_j})/101$, where CNV(i) was the estimated relative copy number, and $E_k(O_j)$ was mRNA level, of the $\mathrm{i^{th}}$ gene in the cell k at the whole genomic scope. 8) The differentially expressed genes (DEG) between tumor-associated monocyte/macrophages C1 and C0 were summarized with the "Findmarker" Seurant package. Setting $|\log_2$fold Change$|$ >1.3 and FDR<0.05 as the cutoff criteria, the $\mathrm{Log_2}$Fold Change >1.3 was regarded as the characteristics of the genes for the early differentiation of TAMM, while $\mathrm{Log_2}$FoldChange<-1.3 as the characteristics of the genes for the late differentiation of TAMM (**Supplementary Table S4**). For GSE103322 cohort: 1) PCA analysis was performed on the data integrated by "SCT" Seurat package. According to the specific markers, the cells were divided into the malignant epithelial cells (KRT14, KRT6A, EPCAM, n = 1939), Cancer associated fibroblasts (FAP, PDRN, n = 1,697), T cells (CD2, CD3D, n = 1,633), B cells (SLAMF7, CD79A n = 354), endothelial cells (PECAM1, VWF n = 75), and mono-macrophage cells (CD14, CD163, CD68, n = 146). 2) T cells were further classified with Multimodal reference mapping into eight clusters of CD4 CTL, CD4 Navie cells, CD4 TCM, CD4 TEM, CD8 Navie cells, CD8 TCM, CD8 TEM, and proliferating T cells (**Supplementary Table S5**).

CIBERSORT and ESTIMATE for Immune Cell and Stromal Scores

For the one TCGA HNSCC and five GEO HNSCC cohorts, "CIBERSORT" and "ESTIMATE" R package were applied to calculate the contents of the 22 kinds of immune cells (1,000 permutations) and immune and stromal score.

The Unsupervised Clustering on TCGA Squamous Cell Carcinoma of Head and Neck and GSE65858 Cohorts

According to the scores of the genes in the early TAMM differentiation and the 22 kinds of immune cells, the unsupervised clustering (through "ConsensuClusterPlus" R package) was applied to the samples with Euclidean distance and Ward (unsquared distances) linkage to get the sub-populations with different immune phenotype and prognosis.

Associations of TCGA Squamous Cell Carcinoma of Head and Neck Subtype With DNA Methylation, CNVs and Mutations

The data of methylation probes were normalized with "wateRmelon" R package, and the difference in methylation probes were analyzed with "limma" R package. The evidently altered regions in genome were screened with GISTIC2.0. The numeric focal CNV values larger than 0.2 meant gain, while less than 0.2 meant loss. Through Somatic mutation data, the TMB (the number of non-synonymous mutations in every million bases of somatic cells) of each patient in the TCGA HNSCC subtype were calculated.

Search for the DEGs in the Subtype of TCGA HNSCS and GSE65858 Cohorts

The DEGs were obtained by comparing the A3 to A1 subtype, and the A3 to B subtype in TCGA HNSCC and GSE65858 cohorts with "limma" in R package. The TCGA HNSCC cohorts were produced by RNA-seq, while the GSE65858 resulted from micro-array. One criterion failed to satisfy the cohorts from a different sequencing approach. If the threshold of GSE65858 was identical to that for TCGA HNSC, the DEGs would be rare. For the TCGA HNSCC cohort, $|Log_2Fold\ Change|>1$ and $FDR<0.05$ were set up as standard. According to the DEGs in GSE65858 array, a threshold of $|Log_2Fold\ Change|>0.2$ and $FDR<0.05$ was selected to keep the numbers of DEGs in the two cohorts from varying too much. FDR was the *p* value calibrated using the Benjamini–Hochberg method.

Confirmation of the Key Genes

In the two HNSCC cohorts, the comparison between A3 and B subtypes gave rise to 181 overlapped candidate genes. Centiscape was applied to the analysis on the protein crosstalk network and was constructed with PPI database. Each knot in the PPI network constructed with 181 DEGs was evaluated with the centiscape of cytoscape for the topo-characteristics, namely, Degree, Eigenvector Centrality, and Betweennesss.

Degree was the most direct and classical index evaluating the regulatory and importance of knot, which was defined as the nodes directly connected to a given node.

Eccentricity $C_{ecc}(v)$ represent the reciprocal inverse of the longest path between the knot v and all other knots. The eccentricity of a node in a biological network can be interpreted as easiness of a protein to be functionally influenced by all other proteins in the same network.

$C_{ecc}(v) = \frac{1}{max\{dist(v,w):\ w \in V\}}$, in which v and w were the nodes in network (V)

S.-P. Betweenness $C_{spb}(v)$ represents the ratio of the path number connecting the knot s and t through v to the total number of path. A high S.-P. Betweenness score meant that the node, for certain paths, was crucial to maintain node connections.

$C_{spb}(v) = \sum_{s\neq v\in V}\sum_{t\neq v\in V}\frac{\sigma_{st}(v)}{\sigma_{st}}$, in which s, t and v were nodes in network (V)

Our purpose was to find out the knot with the higher values on the topo-characteristics, because the higher the value the more significant it was. Since the relative significance of Degree was higher than Eigenvector Centrality, and the Eigenvector Centrality equaled Betweenness, we selected the first 40 knots with the higher Degree. Then, we selected the first 20 DEGs with the higher Eigenvector centrality and Betweenness, respectively. The thresholds of 40 and 20 were set empirically, and had no effect on the outcomes, because the key knots with the higher values of the topo-characteristics would vary with the threshold. According to the descending order of Degree, the first 40 candidate genes were selected. According to the descending sequence of Eigenvector Centrality and Betweenness, the first 20 genes were selected from the 40 candidates (**Supplementary Table S6**). Finally, 12 genes included in both above populations were set up as the hub genes. KEGG database was applied for pathway correlation, CluoGO for visulization, and all the manipulations were based on Cytoscape. From the DEGs by comparing the A3 to A1 subtype of the two HNSCC cohorts, 41 genes were selected for the protein crosstalk network constructed with PPI database.

Pathway Enrichment Analysis

The DEGs from the comparison between the A3 and B subtype in both the TCGA HNSCC and GEO65858 cohorts were applied for GSEA enrichment. Then, the Enrichment map was visualized and annotated. Sample Gene Set Enrichment analysis (ssGSEA) was performed on the TCGA HNSCC, GSE65858, GSE39366, GSE117973, GSE40774, and GSE41613 cohorts with "GSVA" R package to grade the 29 immune signatures (He et al., 2018).

$$\mathrm{Fold-Change} = \frac{1}{\mathrm{n1}}\sum_{\mathrm{i\in immune\ hot}}\mathrm{immune\ related\ score_i} - \frac{1}{\mathrm{n2}}\sum_{\mathrm{i\in immune\ cold}}\mathrm{immune\ related\ score_i}$$

Where n1 and n2 were the number of immune hot and immune cold samples, respectively. Immune related score was the sum of 22 immune cells scores obtained by Cibersort and 29 immune signature scores obtained by GSVA.

The GSVA Scores of the Four Chemokines and the Confirmation of the Immune Hot and Immune Cold Subtype

According to the mRNA levels of the four chemokines, CXCL9, CXCL10, CXCL11, and CCL5, TCGA HNSCC, GSE65858, GSE39366, GSE117973, GSE40774, and GSE41613 cohorts were applied for ssGSEA with GSVA in R package and divided by the median grade into the high- and low-graded subtype, namely, the immune hot and immune cold subtype.

Survival Analysis

The survival curve was generated by "Survminer" R package. The statistical differences among the immune subtype of TCGA HNSCC and GSE65858 cohorts were obtained by log rank test.

Construction of Neural Network

By "neurnet" R package, a neural network containing an input layer, two hiding layers (there were 20 neurons in the first layer, and five neurons in the second layer. Both layers were in dropout), and an output layer.

Activation Function:

$\frac{1}{1+e^{-x}}$; Loss Function: $\frac{1}{N}\sum_i -(y_i * \log(p_i) + (1 - y_i) * \log(1 - p_i))$

Statistical Analysis

All the statistical analyses were performed with R software version 4.0.4. The *t* test and Wilcoxon test were applied for the comparison between two subtypes, while ANOVA was used for comparison among more than two subtype. Fisher exact test was applied for the classified variations between and among subtypes. Pearson or Spearman coefficients were applied for the relevance between two variations. All statistical tests were two-sided and when $p < 0.05$, the difference was regarded as significant.

RESULTS

The Heterogeneity of Tumor-Associated Monocyte/Macrophage in Squamous Cell Carcinoma of Head and Neck

A schematic diagram of the study design and principal findings is shown in **Supplementary Figure S1**. To classify the TAMM in HNSCC according to their differentiation status, 19,718 genes were selected from 39,711 leukocytes of HNSCC patients (**Supplementary Figures S2A, S2B**) qualified for dimensionality reduction with PCA and UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction). The cluster classification analysis with K-NN gave rise to four clusters, which were annotated as NK cells (n = 14,925), T cells (n = 14,073), B cells (n = 2,789), and tumor-associated monocyte/macrophages cells (TAMM; n = 7,924) by SingleR. The 7924 TAMM were further classified with K-NN into seven clusters, in which the cluster 0, 1, 2, and 4 were annotated by SingleR as tumor-associated monocyte/macrophages cells (n = 6,875; **Figure 1A**), while the cluster 3 (n = 312), 5 (n = 478), and 6 (n = 259) as T cells and B cells (data not shown). In the TAMM clusters, TAMMC0, TAMMC1, and TAMMC2 were regarded as TAMM because of the higher CD68 expression, while the TAMMC4 with the lower CD68 expression was considered as dendritic cells (**Figure 1B**). Furthermore, the mature TAM-related genes, such as CD206, CD81 (marker of macrophage M2), TSPO, HLA-DRA, IRF (marker of macrophage M1), and METTL14 (C1q+), were mainly expressed in TAMMC0 and TAMMC2 (the expression in TAMMC0 was higher than that in TAMMC2), but almost silenced in TAMMC1 (**Figure 1C**), indicating TAMMC0 as the mature TAM, TAMMC1 as the early tumor-associated monocytes (TAM-M0), and TAMMC2 as the transforming TAM-M0 from monocytes to macrophages. Differential gene expression analysis between TAMMC1 and TAMMC0 classified 54 genes highly activated in TAMMC1, including S100A12, S100A8, VCAN, PTGS2, and CD55, into the early group of TAMM differentiation, and the other 51 genes robustly expressed in TAMMC0, such as C1QB, C1QC, MMP12, and SPP1, into the late group. Such a difference was also proven by pseudotime clustering heat map (**Figure 1D**, **Supplementary Figure S1C**). By analyzing the scRNA-seq data with GSVA and CIBERSORT, the gene function in TAMMC1 was enriched in cellular toxicity and immune inflammation, as well as the stemness and metabolism. In contrast, the gene function in TAMMC0 was less enriched in stemness and metabolism, but more highly enriched in cellular toxicity and immune inflammation, as well as the pathways of hypoxia and angiogenesis. The enriched gene function of TAMMC2 was medially located between TAMMC0 and TAMMC1 (**Figure 1D**; **Supplementary Figure S3**). Therefore, TAMMC1 was highly scored as early tumor-associated monocyte/macrophages and TAMMC0 as mature macrophages.

Differentiation Trajectory and Copy Number Variation Verified the Heterogeneity of Tumor-Associated Monocyte/Macrophage

According to the above differential gene expression, the differentiation trajectory of TAMM was established, in which TAMMC1 was located in the early stage, TAMMC0 in the late stage, and TAMMC2 diffusely distributed in the early and late stages (**Figure 2A**). Along with the time progression, TAMMC1 was decreased with the increase of TAMMC0, while TAMMC2 was increased and then decreased in the diffusion maps (**Figures 2B,C**). In the cell communication network, the centrally located TAMMC2 exhibited a strong connection with both TAMMC0 and TAMMC1 (**Figure 2D**), which coincided with the finding that TAMMC2 was located medially between TAMMC0 and TAMMC1 in the differentiation trajectory. Similarly, CNV assay revealed that the copy number and deficiency in TAMMC1 genome were relatively lower compared to those in TAMMC0 and TAMMC2 (**Figure 2E**). Therefore, in the heterogeneous TAMM subpopulations of HNSCC, both the gene expression profile and genomic properties indicated that TAMMC0 represented the mature TAM, TAMMC1 stood for the early differentiating monocytes, and TAMMC2 was the monocytes transforming into macrophages.

A Criterion Classifying Squamous Cell Carcinoma of Head and Neck With Different Immune Phenotypes and Prognosis by Combining Tumor-Associated Monocyte/Macrophage Differentiation and Immune cells

As mentioned above, to explore the heterogeneity of TAMM in HNSCC patients, we identified 54 genes as the signature of early

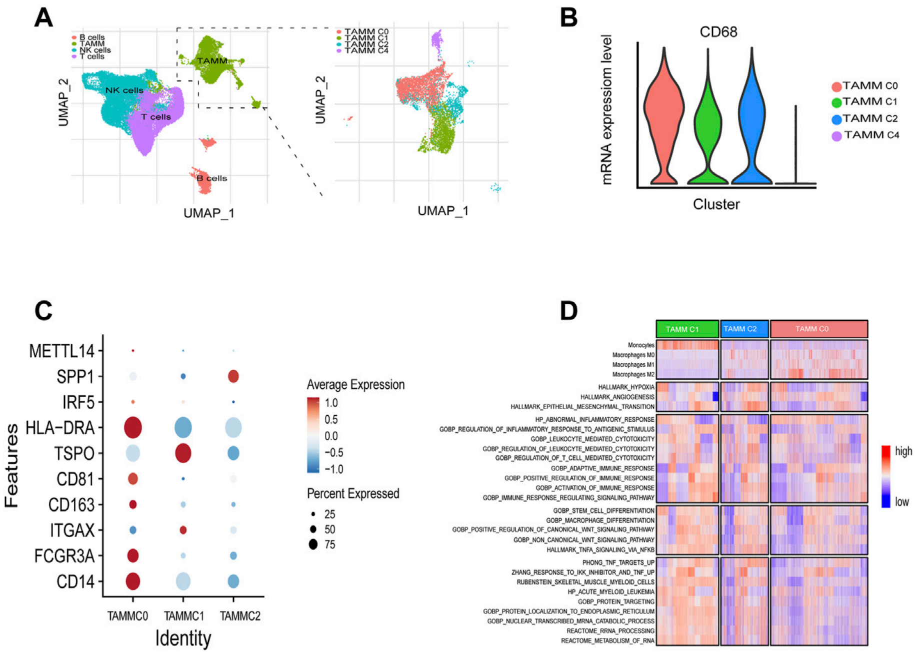


FIGURE 1 | Cellular heterogeneity of tumor-associated monocyte/macrophages/macrophages at the single cell level. **(A)** Umap plots of all clusters annotated by SingleR. **(B)** CD68 expressions levels of TAMM sub-clusters (C0, C1, C2, C4). **(C)** CD14, CD16, ITGAX, CD201, CD81, TSPO, HLA-DRA, IRF5, SPP1, and METT14 expressions levels of TAMM sub-clusters (C0, C1, C2). **(D)** GSVA revealed the enrichment scores of TAMM sub-clusters in the pathways of tumor invasion, immunity, stemness, and metabolism.

TAMM differentiation, and another 51 genes as the signature of late TAMM differentiation. To disclose the correlation of TAMM differentiation with the immune phenotypes of HNSCC, we combined the differentiation signatures with 22 immune cells to form a two-step classifier. First, the scores of 22 immune cells estimated by CIBERSORT were applied for the unsupervised clustering. Both the TCGA HNSCC and the GSE65858 cohorts were classified into A and B subtypes (TCGA HNSCC A = 265, B = 138; GSE56858 A = 175, B = 95) (**Supplementary Figures S4A, S4D**). The PD-1L and IFNG expression was higher in the A subtype than those in B subtype in both cohorts ($p < 0.001$, GSE65858: PD-1L $p < 0.1$). Second, unsupervised clustering was performed in the A subtype with the 54 genes as the early TAMM differentiation signatures and the 51 genes as the late TAMM differentiation signatures. The unsupervised clustering with the early TAMM differentiation signatures could divide A subtype into three subtypes (TCGA HNSCC A1 = 40, A2 = 96, A3 = 127; GSE65858 A1 = 32, A2 = 74, A3 = 69) with different clinical outcomes and immune signatures (**Supplementary Figures S4B, S4C, S4E, S4F**). Interestingly, the three subtypes from A subtype also showed the distinct immune infiltration and immune excluded signatures. Both the immune and stromal scores of the A1 and A3 subtypes were significantly increased compared to those in A2 and B subtypes (**Supplementary Figure S5A**). Moreover, the A1 subtype exhibited the stronger immune infiltration and immune excluded signatures, the A2 subtypes in both cohorts displayed the weaker immune infiltration and immune excluded signatures, while both the A3 subtypes possessed the stronger immune infiltration signatures and the weaker immune excluded signatures. In contrast to A subtype, the B subtype were weaker in immune infiltration signatures and stronger in immune excluded signatures (**Figures 3A,B,F,G**). The PCA with the 54 early TAMM differentiation signatures also supported this notion (**Figures 3C,H**). On the other hand, the unsupervised clustering with the late TAMM differentiation signatures failed to distinguish the immune phenotypes of HNSCC (data not shown). Thus, the A3 subtypes were defined as the high immune infiltration type, and the B subtype as the high immune evasion type. The following survival assay revealed the different prognoses among the subtypes, especially between A3 and B subtype ($p < 0.05$; **Figures 3D,I**). According to TMN staging, the A3 subtypes of both cohorts exhibited a lower ratio of IV stage patients compared to other subtypes (**Figures 3E,J**). To further verify the correlation between TAMM differentiation and immune phenotypes, we performed GSEA analysis on the differentially expressed genes between the A3 subtype and B subtype. The genes highly expressed in the A3 subtype were enriched in immune-associated pathways, such as activation of immune cells, adherence, proliferation, immune response, and regulation, while the genes highly expressed in the B subtype were enriched in cellular development and ECM-related pathways, for instance, mesenchymal development, pattern

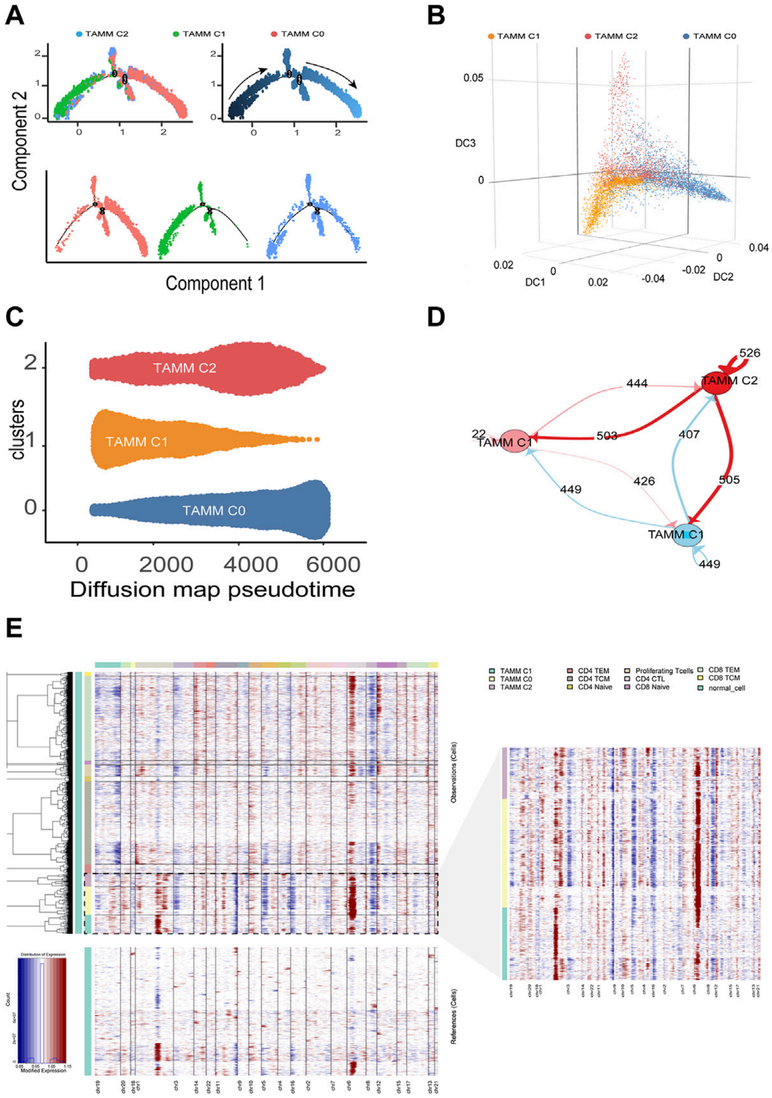


FIGURE 2 | Differentiation trajectory and CNV changes of TAMM. **(A)** Monocle2 reveals the differentiation trajectory of TAMM. **(B)** Three-dimensional diffusion map embedding of macrophages reveals the different differentiation states of TAMM sub-clusters. **(C)** Density diffusion maps model revealed the content of sub-clusters of TAMM at pseudo-time. **(D)** Cell-Cell interaction network of different TAMM sub-clusters, node represent TAMM sub-cluster, and the number of lines represent ligand interactions between two sub-clusters. **(E)** Heatmap of the inferred CNV in which genes were sorted by genomic location.

formation, and cytodifferentiation (**Figures 3K,L**). These results suggested that the early differentiation signatures of TAMM were associated with the HNSCC immune phenotypes and prognosis.

Multiomic Characteristics Associated With the Immune Phenotypes of the Different Squamous Cell Carcinoma of Head and Neck Subtypes

Finally, CNV, SNP, and methylation levels were examined to further explore the immune phenotype in the subtype of TCGA HNSCC cohort. It was found that the mutation frequency of tumor mutation loading and tumor driver genes (TP53, TTN, etc.) in the A2 subtype was noticeably higher than that in other subtypes (**Supplementary Figures S5B, S5C**). CNV analysis found that the focal copy numbers in 3p, 11q, and 2p were significantly distinguishable between the A3 subtype and B subtype (**Supplementary Figures S5D–E**). The methylation assay revealed that there were 96 genes highly expressed in B subtype overlapped with the methylation probe highly expressed in A3 subtype, while only 13 genes highly expressed in A3 subtype were detected by the methylation probes highly expressed in B subtype (**Supplementary Figures S5F–G**). These findings implicated that CNV, SNP, and methylation levels also contributed to the different immune phenotypes in the HNSCC subtypes.

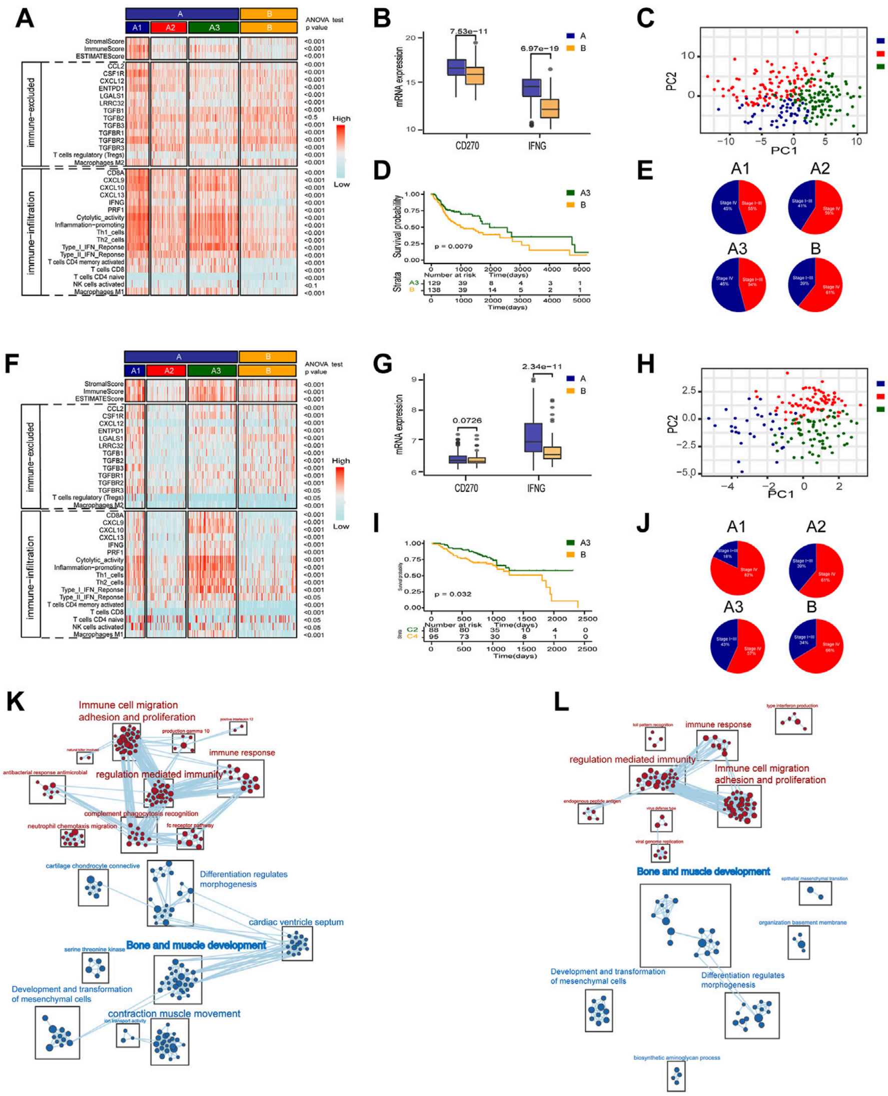


FIGURE 3 | A two-step molecular classification combining the early differentiation features of TAMM and 22 immune cell scores. **(A,F)** Heatmaps of immune-related components, stromal score, and tumor purity score of the HNSCC subtypes in TCGA HNSCC **(A)** and GSE65858 cohorts **(F)**. **(B,G)** IFNG, PD-L1 expression level of subtype of TCGA HNSCC **(B)** and GSE65858 **(G)**. **(C,H)** PCA of the mRNA expression of 54 early differentiation feature genes from the HNSCC patients in the TCGA **(C)** and GSE65858 cohorts **(H)**. **(D,I)** Kaplan-Meier curves for overall survival (OS) of all HNSCC patients in TCGA **(D)** and GSE65858 **(I)** within A3 and B subtypes. **(E,J)** The pie chart showed the proportion of TMN stages with four different immunophenotypes in TCGA **(E)** and GSE65858 cohorts **(J)**. **(K,L)** GSEA network of DEGs in A3 vs B subtypes using Enrichment map in TCGA **(K)** and GSE65858 cohorts **(L)**.

Construction of Gene Regulatory Network Based on Differential Genes Between Squamous Cell Carcinoma of Head and Neck Subtype

To explore the mechanisms regulating the formation of different subtypes, we compared the differential expressed genes (DEGs) between A3 and B subtypes, and between A3 and A1 subtypes. There were 451 highly DEGs in the A3 subtype compared to the B subtype in the TCGA HNSCC cohort (**Figure 4A**), and 567 highly DEGs in the A3 subtype compared to the B subtype in GSE65858 cohort (**Figure 4B**). There were 181 overlapped genes in the two groups of the highly DEGs (**Figure 4C**), which represented the high immune infiltration associated genes in HNSCC. On the other hand, we obtained 659 lowly DEGs in the A3 subtype from the comparison to the A1 subtype of TCGA HNSCC cohort (**Figure 4D**), and 596 lowly DEGs in the A3 subtype from the comparison to the A1 subtype of GSE65858 cohort (**Figure 4E**). In the two groups of lowly DEGs, 41 genes were overlapped (**Figure 4F**). Thus, the high immune infiltration-associated genes overlapped evidently between different cohorts, while the immune evasion-related genes showed diversity between different cohorts even in the instance of high immune infiltration. By exploiting STRING database, the 181 highly and 41 lowly DEGs were constructed into a protein crosstalk net (**Figure 4G**). Moreover, in the 41 immune evasion-related genes, those correlated with the genes

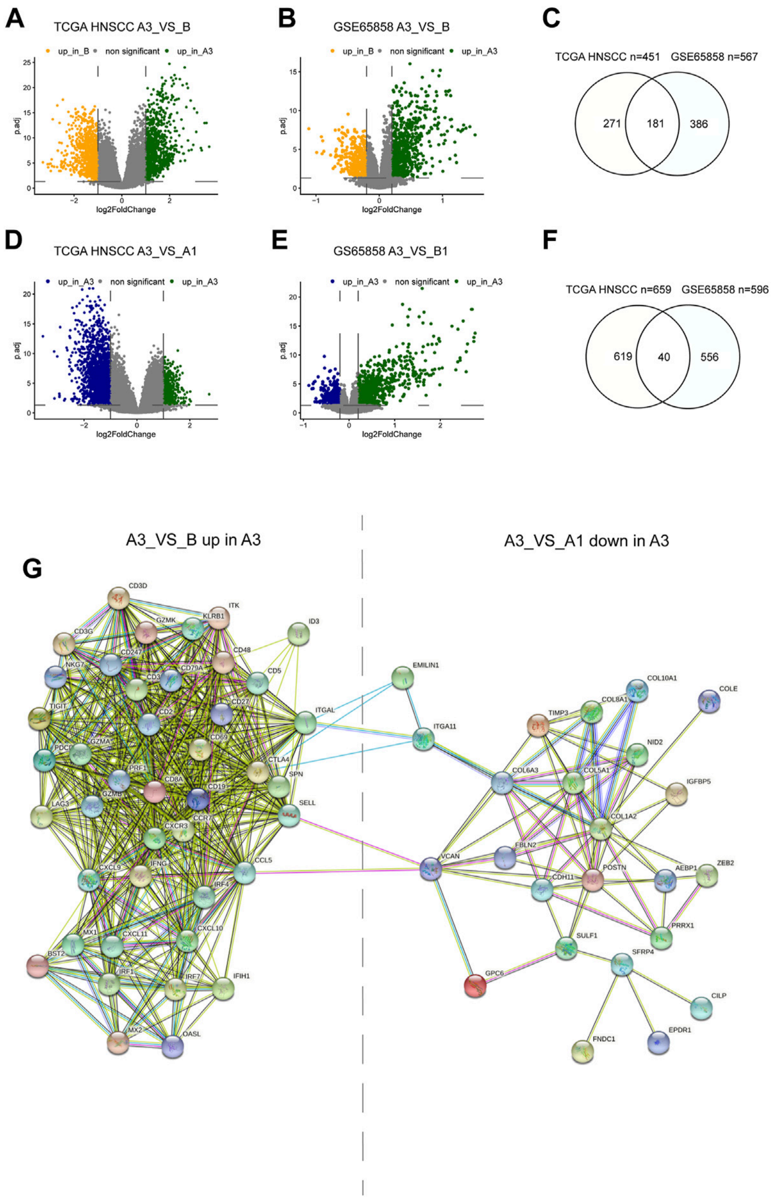


FIGURE 4 | Differential genes between subtypes and protein regulatory network. **(A,B,C)** Volcano plot and Venn diagram show DEGs of A3 vs B in TCGA HNSCC and GSE65858 cohorts. **(D,E,G)** Volcano plot and Venn diagram showed the DEGs of A3 vs A1 in TCGA HNSCC and GSE65858 cohorts. **(F)** PPI protein regulatory network of 181 overlapping DEGs up in A3 (A3 vs **(B)** and the 40 overlapping DEGs down in A3 (A3 vs A1).

encoding extracellular matrix (POSTN, COL6A3, COL1A2, etc.) resided in the core of the network.

Tumor-Associated Monocyte/Macrophage-Associated Chemokines-CXCL9, CXCL10, and CXCL11-and Inflammatory Chemokine-CCLL5 Were Key Nodes in Gene Regulatory Network

To screen out the key nodes in the gene regulatory network of the high immune infiltration we assumed three criteria: at the center of the regulatory network, belong to the same pathway, and highly correlated expression. In the network constituted by the 181 highly DEGs, we screened out 40 highly regulated genes with the Degree more than 30. Although the correlation matrix also verified the high association among the 40 highly regulated genes (**Supplementary Figure S6A**), the Degree and correlation are insufficient for the identity of the key genes. Thus, the 40 candidate genes were arranged in the order of Betweenness which represented the center value of the node (**Figure 5A**) and Eigenvector according to the importance of integrating adjacent nodes (**Figure 5B**), respectively. Then, by comparing the first 20 genes arranged with Betweenness to the first 20 genes arranged with Eigenvector, 12 overlapped genes were chosen as the hub genes (**Figure 5C**). Finally, the function of these 12 hub genes were applied for KEGG pathway correlation with CluoGO (**Figure 5D**) through which four chemokines (CCL5, CXCL9, CXCL10, and CXCL11) were screened out. Although not included in the 12 hub genes, CXCL11 shared the same family with CXCL9 and CXCL10, and was highly correlated with their expression levels. So CXCL11 was also identified as one of the driver genes. These four chemokines were regarded as the key node in gene regulatory network of the high immune infiltration, because of these characteristics: 1) the core genes with the higher Degree, Betweenness, and Eigenvector in the network; 2) robust relevance at the transcription level (**Supplementary Figures S6B–D**), and the remarkably higher difference in CXCL9, CXCL10, and CXCL11 (A3 vs B) than the other eight genes (**Supplementary Figure S6E**); and 3) belong to the same pathway and share the higher topological signs. Moreover, during TAMM differentiation, CXCL9, CXCL10, and CXCL11 were increased with the time progression in Pseudotime analysis (**Figure 5E**). Taken together, the four chemokines with the strongest functional co-regulation and co-expression could be regarded as the pivotal genes screening the high immune evasion of HNSCC.

The Transcription of the four Chemokines-CXCL9, CXCL10, CXCL11, and CCLL5-Was Influenced by Epigenetic and Health Factors

To further explore the endogenous (epigenetic) and exogenous (health manner) factors impacting the expression of the four chemokines-CXCL9, CXCL10, CXCL11, and CCL5--the chromatin accessibility and the methylation status at the transcription initiation regions of CXCL9, CXCL10, CXCL11, and CCL5 were analyzed in the whole genome with TCGA HNSCC ATAC and methylation database. The enriched reads were evidently concentrated at the transcription initiation regions of CXCL9, CXCL10, CXCL11, and CCL5 in the A3 subtype, compared to the A1 and A2 subtypes (**Figure 5F**), implicating a more active transcription of CXCL9, CXCL10, CXCL11, and CCL5 in the A3 subtype. In contrast, the methylation levels of CXCL9, CXCL10, CXCL11, and CCL5 showed insignificant difference among the subtype, implying that the four chemokines were epigenetically regulated by the manners rather than DNA methylation. However, the methylation of CXCL9, CXCL10, CXCL11, and CCL5 in the control patients were higher (**Figure 5G**), suggesting de-methylation of the four chemokines was crucial for HNSCC genesis. We also found that the transcription and methylation levels of CXCL9, CXCL10, CXCL11, and CCL5 were correlated with age, smoking, alcohol consumption, and HPV infection. The higher mRNA levels of the four chemokines were detected in the HNSCC population with older age and lower consumption of tobacco and alcohol (**Supplementary Figure S7A, S7C, S7E**; *t* test, $p < 0.05$). The HPV positive HNSCC group exhibited a higher CXCL10 mRNA level and an increased methylation of CXCL9 and CXCL11 compared with the HPV negative HNSCC group (**Supplementary Figures S7B, S7D**; *t* test, $p < 0.05$). Thus, it was concluded that, although not associated with the immune phenotypes of HNSCC, the transcription of the four chemokines regulated by DNA methylation and health factors were critical for HNSCC genesis.

The Four Chemokines-CXCL9, CXCL10, CXCL11, and CCL5-Positively Regulated Immune Responses and Were Associated With the Low CNV and CASP8 Mutations in the Squamous Cell Carcinoma of Head and Neck Genome

According to the mRNA levels of the four chemokines, six HNSCC cohorts (one TCGA cohort and five GEO cohorts) were graded with GSVA, and then divided into the high- and low-graded groups with the median grade to evaluate the correlation of the four chemokines with immune response. There was a remarkable difference between the high- and low-graded groups in the immune signature and genome. The TCGA and most high-graded cohorts showed a higher enrichment of immune cells (Macrophages M1, CD4 T cells memory activated, and CD8 T cell) in the infiltration grading of the 22 kinds of the immune cells, got higher scores in the enrichment of antigen present during tumor immune circle, immune cell infiltration, and the recognizing and killing of tumor cells by effector T cells in the 29 immune signatures assay (**Figure 6A**), and was given the lower scores in the TIDE assay. All of the results suggested a better response to immune therapy and was verified by the cohorts of immune therapy, in which the GSVA grades of the four chemokines in the CR group were higher than those in PR, SD, and PD groups (**Figure 6B**). Based on these findings, we classified HNSCC into the immune hot and immune cold phenotype according to the GSVA grades of the four chemokines. Then, we estimated the distribution of the mutations from the first 30 HNSCC driver genes with the highest frequency of mutation (TP53, TTN, CSMD3, SYNE1, etc.,) in the hot and cold immune groups, and found that except for CASP8, all other driver genes had an elevated frequency of mutation in the low-graded group (**Figure 6C**; **Supplementary Table S7**), implying that the mutations of CASP8 endowed HNSCC with a stronger immunity. Moreover, CNV analysis revealed that in the immune cold group, an active CNV was detected in several hot spot regions (gain: 3p, 8q, 17q, 18p. loss: 2q, 7q, 13q) (**Figure 6D**). In combination with Kech classification, we found that the most immune hot was BA type, while the most immune cold was CL type (**Figure 6E**). All the above results suggested that the four chemokines could not only act as the markers identifying the HNSCC with high concentration of immune cells (CD8 T cell, Macrophage M1, etc.), but also reflect the HNSCC characteristics comprehensively. It was also suggested that the immune hot subtype of HNSCC could enhance the immune responses through CASP8 mutations, and the immune cold subtype also circumvented immune responses through gene mutations. Further exploration on the crosstalk among the four chemokines, TME, and immune cells in the immune circle by analyzing the relevance in TCGA cohort disclosed that CXCL9, CXCL10, CXCL11, and CCL5 showed a strongly

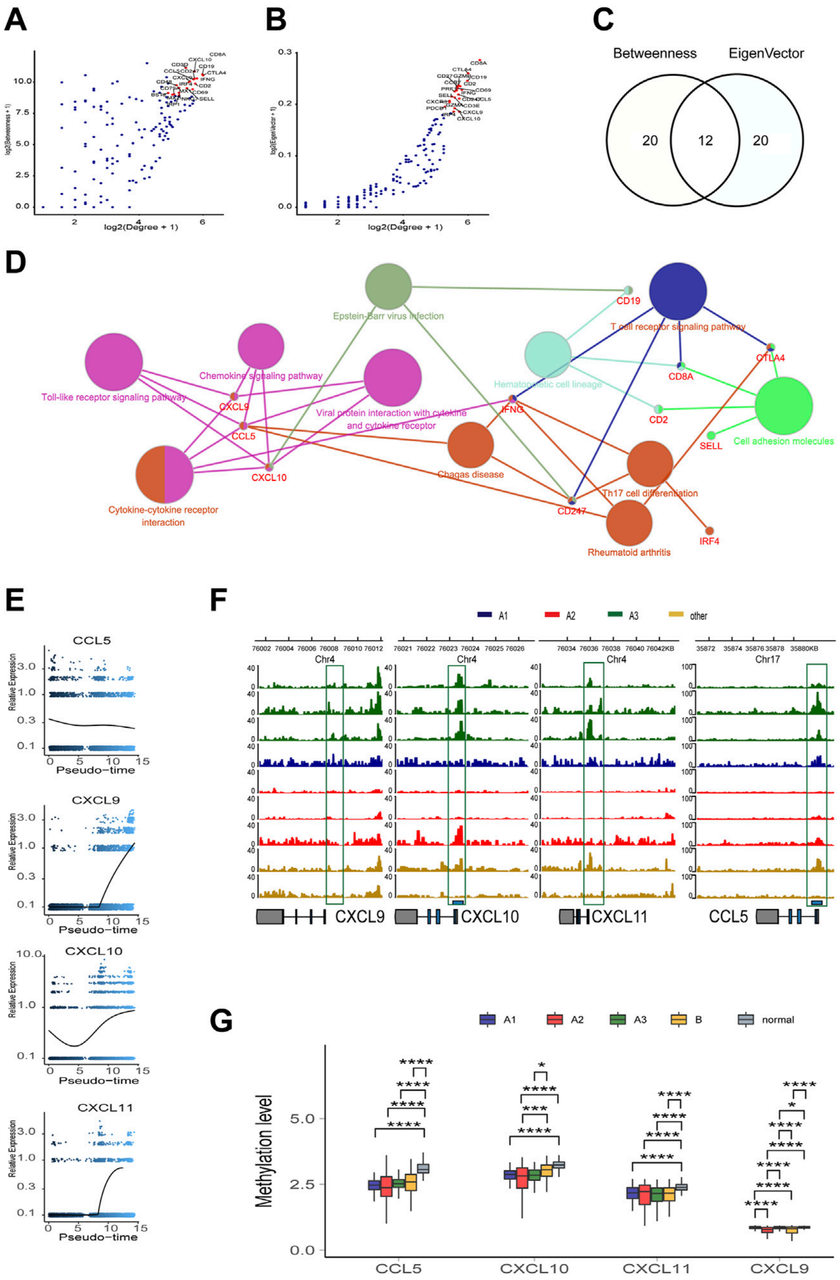


FIGURE 5 | The topological feature of the network composed of 181 genes. **(A)** Dotplot of -log2(Degree+1) and log2(Betweenness+1) in selected 181 nodes. **(B)** Dotplot of log2(Degree+1) and log2(Eigenvector+1) in selected 181 nodes. **(C)** Venn diagram of top 20 genes in A or **(B) (D)** KEGG pathway association network of 12 hub genes. **(E)** Trend of mRNA expression levels of the four chemokines (CXCL9,10,11, and CCL5) following the differentiation trajectory of TAMM. **(F)** Aggregation of ATAC-seq peaks of the four chemokines of nine patients in the transcription initiation region. **(G)** Methylation levels of the four chemokines in different HNSCC subtypes.

positive association with immune cells (Macrophages M1, CD4 T cells memory activate and CD8 T cells), and the three stages of tumor immune circle (**Figure 6F**). Since the similar association was also detected in other HNSCC cohorts, the four chemokines were proven to enhance the anti-tumor immune capability. Moreover, we also found that macrophages M1 was strongly positively associated with the activation of dormant CD8 and CD4 T cells (**Figure 6G**).

The Relationship Between the Four Chemokines and the Sub-Populations in the Squamous Cell Carcinoma of Head and Neck Tumor Micro-Environment at Single Cell Level

The cohorts of GSE10332 and GSE139324 (the GSE10332 cohort contained all kinds of cells in TME, while the GSE139324 cohort

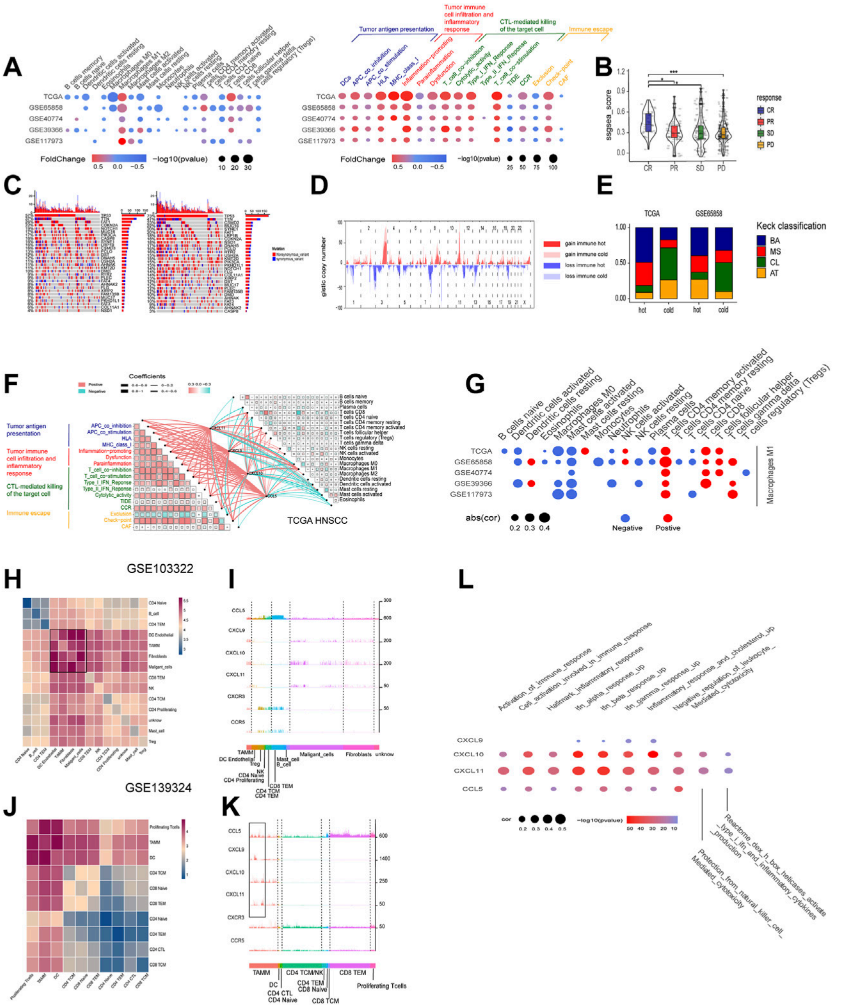


FIGURE 6 | The roles of the four chemokines (CXCL9,10,11, and CCL5) in HNSCC TME and their association with tumor genomic changes. **(A)** Dotplot summarized the scores of 22 immune cells estimated by GIBESORTRT and GSVA scoring on the fold change and *p*.adjust of 29 immune signature between the immune hot and cold group classified by the median GSVA of the four chemokines (only $p < 0.05$ was given). **(B)** Vilionplot summarized the GSVA scores of CXCL9, CXCL10, CXCL11, and CCL5 of four outcomes (CR, PR, PD, SD) in the cohort receiving anti-PD-1/PD-L1 immunotherapy. **(C)** The OncoPrint was constructed between high and low scores of the top 30 genes with the highest mutation frequency. **(D)** CNV plot showed the frequency of copy-number gains (red) and deletions (blue) among immune hot and cold groups of the TCGA-HNSC cohort. **(E)** The stacking histogram showed the distribution of Keck classification in the immune hot and cold groups of TCGA HNSCC and gse65858 cohorts. **(F)** Correlation between the four chemokines and GIBESORTRT score of 22 immune cells in HNSCC-TCGA and 29 immune signature GSVA score. The cell charts in the upper-right triangular exhibited the correlation among the 22 immune cell scores, and the cell charts in the -lower triangular showed the correlation among the 17 immune signature scores (Red stood for positive and green for negative, the darkness and lightness of the colors for the high and low of the coefficients, and the size of the cell for *p* value). The lines between two cell charts represent the correlation of the mRNA of the four chemokines with the bilateral immune scores (Red stood for positive and green for negative, and the thickness of the lines for the high and low of the coefficients). **(G)** Dotplot summarized the correlation coefficients between the scores of 21 immune cells estimated by GIBESORTRT and the scores of Macrophage M1 estimated by GIBESORTRT (only $p < 0.05$ was given). **(H,J)** Heatmap of average expression level of ligand–receptor interactions in all clusters in GSE103322 and GSE139324. **(I,K)** Histogram of expression levels of CCL5, CXCL9, CXCL10, CXCL11, CXCR3, and CCR5 in each cell in GSE103322 and GSE139324 cohorts. **(L)** Dotplot summarized the correlation coefficients between the expression level of CXCL9, CXCL10, CXCL11, and CCL5, and the GSVA of 10 immune signatures in each cells classified as TAMM.

only contained infiltrative leukocytes) were applied for the relationship assay at the single cell level to disclose the effects of CXCL9, CXCL10, CXCL11, and CCL5 on the HNSCC subpopulations in TME. Cell communication assay with CellPhoneDB found that TAM, DC, tumor-associated fibroblasts (CAF), and malignant epithelium cells communicated with themselves and other cells intensively (**Figures 6H,J**). Interestingly, CXCL9, CXCL10, and CXCL11 were expressed robustly in TAM, malignant epithelium cells, and CAFs, but weakly in CD4 and CD8 T cells. However, CXCR3, the common receptor for CXCL9, CXCL10, and CXCL11, was expressed in CD4 and CD8 T cells

(**Figure 6I**), indicating that CD4 and CD8 T cells were chemoattracted to immune focus by the CXCL9, CXCL10, and CXCL11 emanated from TAM, malignant epithelium cells, and CAF. Moreover, CCL5 and its receptor, CCR5, were mainly expressed in CD4, CD8 T cells, and NK cells (especially CD8 T cells; **Figure 5K**), suggesting that CCL5 influenced T cells through autocrine or paracrine. Worthy of note, the mRNA peaks of CXCL9, CXCL10, CXCL11, and CCL5 in TAM were distributed in multiple sub-populations (**Figure 6K**), suggesting that although the four chemokines were highly correlated at the bulk RNA-seq level, such correlation could be inconsistent at the single cell level due to the heterogeneity of TAMM. Finally, the relevance assay on TAM clusters at the single cell level disclosed that the mRNA levels of the four chemokines were positively correlated to the pathways involved in immune infiltration and immune inflammation response (**Figure 6L**). Therefore, the expression of the four chemokines were associated with the immune sub-populations in TAMM.

The Immune Hot Subtype of Squamous Cell Carcinoma of Head and Neck Characterized by the High Expression of the Four Chemokines was Sensitive to Anti-Cancer Drug Targeting ERK1-MARK and RAS Pathway

The GSEA assay on the DEGs between the high- and low-graded groups of the six HNSCC cohorts revealed that the pathways enriched in the high-graded groups were mainly involved in chemoattraction and immune response of immune cells, as well as cell proliferation and differentiation, which were consistent in different cohorts. In contrast, the pathways enriched in the low-graded groups were relatively sparse in different cohorts, though mainly concentrated in tumorigenesis and progression, such as stemness, metastasis, metabolism, and hypoxia (**Figure 7A**). The relevance assay on the IC50 of the HNSCC cell lines treated with anti-tumor drugs in GDSC database disclosed that the GSVA scores of CXCL9, CXCL10, CXCL11, and CCL5 were negatively correlated to ERK1-MARK (**Figures 7B–H**) and RAS pathway (**Figures 7I,J**) targeted by anti-HNSCC drugs, implicating that the susceptibility of the immune hot subtype of HNSCC to the drug originated from the influence on ERK1-MARK and RAS pathway.

A Neural Network Predicting Model With the Four Chemokines and Three Immune Cells

Through the above studies, we found out a strongly positive correlation between four chemokines (CXCL9, CXCL10, CXCL11, and CCL5) and three kinds of immune cells (Macrophage M1, CD8, and CD4 T cells) in the immune phenotypes of HNSCC. Since the above seven factors were located in the core of the immune response in HNSCC, we designed a 4-layer neural network to predict the response (response: CR, PR. not response: SD, PD) to the treatment with anti-PD-1/PD-L1 (**Figure 8A**), in which GSE154538, GSE141119, GSE91061, GSE78220, GSE176307, and the IMvigor210 acted as the training set (n = 501), and GSE93157 as the test set (n = 65). The AUC of the training set and the test set in the prediction model reached 95% and 74.6%, respectively (**Figure 8B**). The confusion matrix of the training and test sets was displayed in **Figures 8C,D**.

DISCUSSION

Despite the great progress made by ICB in multiple tumors, only a minority of HNSCC patients have benefited from ICB therapy. It is of major importance to characterize the HNSCC sub-population susceptible to ICB therapy, and the key genes maintaining the sub-population (Curran et al., 2010). Although a variety of criteria were proposed previously, few of them concern the association of TAMM with the immune responses in HNSCC. Since TAMM in TME contributed greatly to tumorigenesis and progression (Estko et al., 2015; Gao et al., 2016), a lot of researchers focused on the potentials of TAMM in immune therapy (Coifman et al., 2005). However, most previous studies and strategies ignored the heterogeneity of TAMM, but treated the TAMM as a static entity, which meant the immune therapy was challenged by TAMM heterogeneity. In the present study, we applied bioinformatical methods from multiple dimensions to explore the heterogeneity of the TAMM during differentiation, as well as its correlation with the immune responses in HNSCC. Although CIBEOSORT was not reported to be applied in the matrix of single-cell RNA-Seq, we think that, according to the resolve of non-negative matrix in the algorithm theory of CIBEOSORT, the bulk RNA-seq could be resolved into the matrix of single-cell RNA-Seq timing the matrix of cell type clusters. In our study, the single-cell RNA-Seq matrix, namely, the Leukocyte signature matrix (LM22), was set as the decision variable, and the single sample in the bulk RNA-seq as response variable for SVM linear deconvolution. The outcome weighted vector was regarded as the cell type abundance of each sample. Therefore, it is sound to transform the bulk RNA-seq matrix into Single-cell RNA-Seq matrix by treating each cell as a sample in the bulk RNA-seq. By treating the Single-cell RNA-Seq data with the above algorithm theory, we found that combining the genes characterizing early TAMM differentiation with the immune cells in HNSCC could provide a criterion classifying the immune subtypes of HNSCC more precisely. Furthermore, the four chemokines-CXCL9, CXCL10, CXCL11, and CCL5-were identified not only as the driver genes initiating and maintaining the immune hot subtype of HNSCC, but also the nodes connecting TAM (macrophage M1), CD4, and CD8 T cells together.

The heterogeneity of TAMM in HNSCC was found to result from not only the different subpopulations, but also the different stages during differentiation. Molecularly, the gene expression and genomic constitution of TAMM underwent remarkable alterations during the differentiation from tumor-associated monocyte to tumor-associated macrophages, which was also supported by the *in vitro* assay (Singhal et al., 2019). Previously, a criterion exploiting 22 kinds of immune cells was

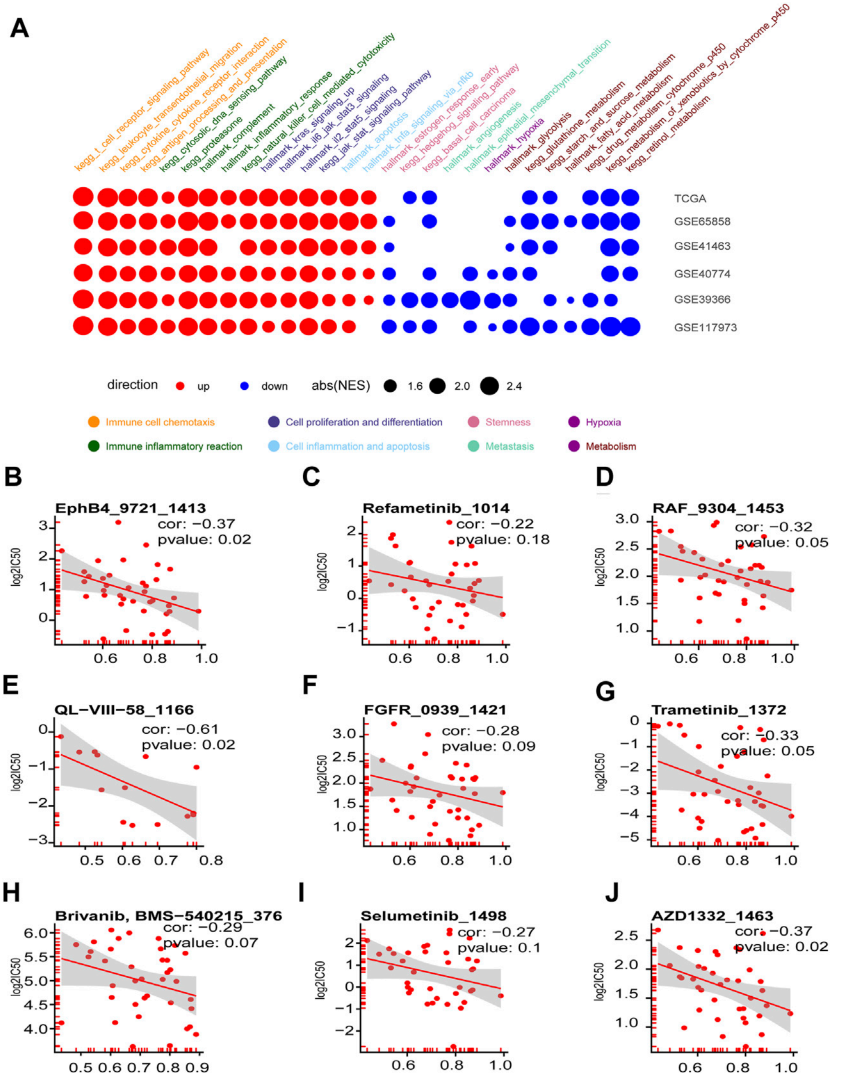


FIGURE 7 | Enriched tumorigenesis pathways in immune hot and cold phenotypes determined by the four chemokines (CXCL9,10,11, and CCL5) and immune hot sensitive drugs. **(A)** Dotplot summarized the GSEA NES scores of signal pathways related to tumor immunity and tumorigenesis between immune hot and cold groups in the six HNSCC cohorts (only *p*.adjust<0.05 were given). **(B–J)** Scatter diagram summarized the correlation coefficients between GSVA score of CXCL9, CXCL10, CXCL11, and CCL5, and the $\log_2$(IC50) in seven drugs targeting ERK-MAPK pathway **(B–I)** and two drugs targeting RAS pathway **(I, J)**.

once proposed to classify HNSCC into the high and low immune enrichment subtype (Feng et al., 2020; Zhang et al., 2020), however, it was too broad to characterize the immune subtype and prognosis. To establish a practical and accurate criterion, we divided the gene expression profile during TAMM differentiation into the early and later groups, and found out the DEGs between the early and later groups. By using multiple bioinformatical methods, the DEGs in the early TAMM differentiation were found, indicating a better prognosis in the subtype with high immune infiltration and low immune evasion (Sanmamed and Chen., 2018). Thus, the DEGs in early TAMM differentiation could reflect more details in the immune infiltration, evasion (which could estimate the immune phenotype more comprehensively) (Turan et al., 2021), and prognosis of HNSCC.

Previous studies reported a correlation of the better prognosis with the elevated tumor mutation burden (TMB) (Chan et al., 2015; Spencer et al., 2016; Büttner et al., 2019). However, in our study, the A2 subtype which had the highest TMB in the TCGA HNSCC cohort displayed the lower immune infiltration and evasion, implicating the insusceptibility to ICB therapy. Since similar characteristics were also detected in another HNSCC clinical cohort (Kim et al., 2020), it was still unclear whether the HNSCC subtype with a higher TMB was susceptible to PD-1/PD-L1 therapy.

The DEGs between A3 and B subtype (high immune infiltration vs high immune evasion) were applied to construct the protein crosstalk network from which the genes highly correlated in expression at the nodes of regulatory network and sharing the same pathway were identified as the driver

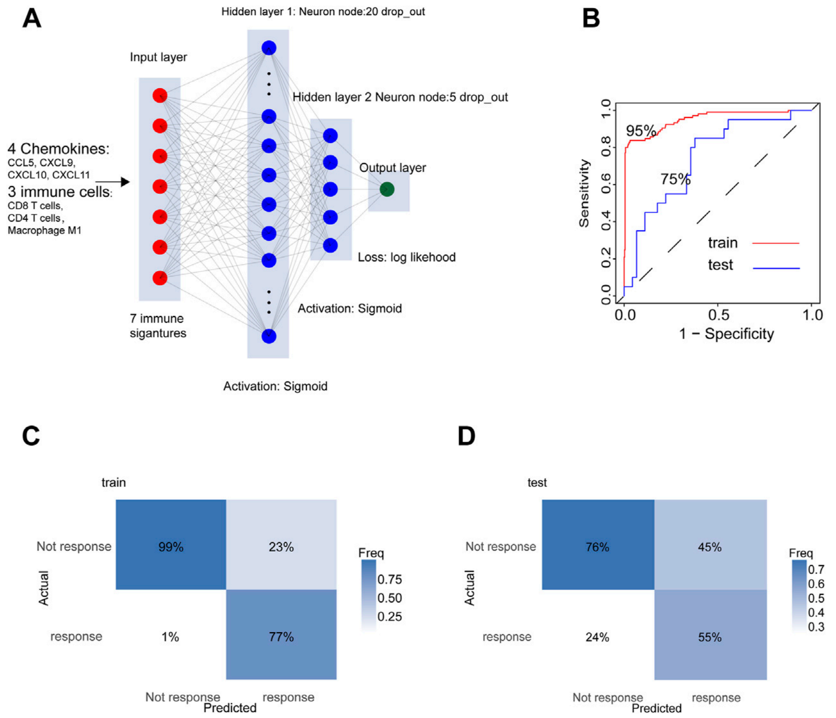


FIGURE 8 | The four chemokines and three immune cells were used to construct a neural network model to predict the response to anti-PD-1/PD-L1 immunotherapy. **(A)** Schematic diagram of the neural network. **(B)** ROC plot of train cohort and test cohort. **(C,D)** The confusion matrix in test cohort and train cohort.

genes initiating and maintaining the immune subtype. The DEGs between the high immune infiltration and high immune evasion (A3 subtype vs B subtype) showed a high similarity in different cohorts. In contrast, the DEGs between high immune infiltration subtype (A1 vs A3 subtype) were different from one another in different cohorts. Therefore, we concentrated on the DEGs between the high immune infiltration and high immune evasion (A3 subtype vs B subtype) to explore the key genes regulating the immune hot phenotype. Since both the A3 and A1 subtypes stood for the high immune infiltration subtypes in the two cohorts, the difference between A3 and A1 subtypes represented the discrepancy between the high immune infiltration subtypes. From **Figure 5G**, it can be seen that such discrepancy could be partially attributed to the various extents of the immune evasion. However, because the number of the intersected DEGs in the A3 and A1 subtypes was too low to contribute to the difference, the factors resulting in the difference were implicated varying dramatically in different cohorts. Since the A3 and B subtypes in the two cohorts exhibited the converse immune signatures, their comparison was supposed to get the core genes associating immune infiltration with the HNSCC. The numerous overlapped DEGs from the two cohorts reflecting the similarity in the high immune infiltration between the two cohorts allowed the following exploration of the pivotal genes in the HNSCC with immune infiltration phenotype. Taking all above findings into account, we concluded that the TAMM differentiation related chemokines-CXCL9, CXCL10, and CXCL11, and inflammatory chemokine CCL5-were the driver genes initiating and maintaining the high immune infiltration phenotype of HNSCC. A series of previous reports supported that the four chemokines could work as the potential targets of immune therapy. During tumorigenesis and progression, the CXCL9-11/CXCR3 axis regulated the differentiation and chemoattraction of T cells (Tokunaga et al., 2018), and the CCL5/CCR5 axis influenced growth and metastasis (Aldinucci et al., 2020). Previous reports showed that CXCL9, CXCL10, and CCL5 could mark T cell–inflamed phenotype of pancreatic cancer (Romero et al., 2020). CXCL9 and CCL5 activated immune responses and enhanced ICB therapy in mouse model of ovary cancer (Dangaj et al., 2019), and the melanoma in CXCR3 knock-out mice exhibited decreased immune infiltration and poor prognosis (Korniejewska et al., 2011). However, we have to acknowledge that there was subjective opinion in the criteria of the pivotal genes. Actually, besides the four chemokines, the other eight candidate hub genes were also verified to act as the pivotal genes establishing and maintaining the high immune infiltration to some extent. However, we think that CXCL9, CXCL10, CXCL11, and CCL5 were the optimal combination, because they satisfied the three criteria: 1). they shared multiple pathways, implicating that they could collectively reflect the activity of immune pathway, instead of independently. Other genes were located in different pathways, which raised the uncertainty for their function, though they possessed the higher network topology; and 2). CXCL9, CXCL10, CXCL11, and CCL5 were highly correlated in expression. We tested the correlated expression of the 181 overlapped DEGs in TCGA cohorts and confirmed a higher correlation among CXCL9, CXCL10, and CXCL11 than among other DEGs. Although CXCL11 was excluded from the 12 candidate genes with the highest topological signatures, it had the higher correlated expression, belonged to the same family and shared multiple pathways with the other three chemokines. Therefore, we selected CXCL9, CXCL10, CXCL11, and CCL5 as the pivotal genes in the high immune infiltration of HNSCC for the following study.

Because the roles of the four chemokines in HNSCC have been little studied, to explore their pivotal effects in the immune responses to HNSCC, we divided the immune circle into infiltration and evasion stages. The immune infiltration included: 1) the release and convey of tumor cell antigen, 2) chemoattraction and infiltration of immune cells into TME by the cytokines and inflammation, and 3) recognition and elimination of HNSCC cells by CTL. With the exhaustion of CTL and the expression of immune suppression factors (TGF-beta, PD-L1, etc.) by HNSCC, the immune evasion commenced. The bulk RNA-seq data revealed that CXCL9, CXCL10, CXCL11, and CCL5 were positively correlated to all three steps of immune infiltration, but negatively to immune evasion. The scRNA-seq data further disclosed that the four chemokines and their receptors were highly expressed in DCs, enhancing the antigen presentation in TME; the high CCL5 expression in CD8 T cells, NK cells, and certain TAM promoted inflammation; the tumor cells highly expressing CXCL9, CXCL10, and CXCL11 attracted the CD8 T cells, NK cells, and TAM, which eliminated tumor cells by perforating cell membrane, digesting through serine proteases and apoptosis via ligand binding (Lee et al., 2018). All of these results supported the role of the four chemokines in initiating and maintaining the immune hot phenotype of HNSCC.

According to the GSVA scores of the four chemokines, the HNSCC cohorts were divided into the subtypes of immune hot and immune cold. The immune cold subtype was characterized by the evident alteration on CNV, which was correlated to tumor invasion and decreased immune responses (Davoli et al., 2017). This finding also implied that the decreased immune responses in immune cold subtype resulted from the imbalanced immune gene expression caused by CNV alteration, as opposed to the dysfunction of single immune gene. The other sign of immune cold subtype was the higher mutation frequency, including the tumor driver genes of TP53, TTN, etc. Reversely, the frequency of CASP8 in immune hot subtype was noticeably higher than that in immune cold subtype. Since several studies demonstrated that the CASP8 mutations characterized the local activation of immune cells and inflammation (Rooney et al., 2015; Tummers and Green., 2017), it suggested that CASP8 mutation was capable of activating immune responses in the immune hot subtype.

Worthy of note, the inflammation resulting from the four chemokines could also increase the tumor invasion. As shown in previous studies, the VEGF-PIK3/AKT pathway activated by CCL5 promoted tumor metastasis by stimulating angiogenesis and ECM remodeling (Karnoub et al., 2007; Wang et al., 2012), and the robust expression of CXCL10 also enhanced gastric cancer invasion and metastasis by binding the receptor CXCR3A (Yang et al., 2016). Thus, further exploration was still required to elucidate the relationship between the four chemokines and HNSCC prognosis. Additionally, the four chemokines were strongly associated with the check points on HNSCC surface, implicating that HNSCC cells could circumvent immune attacks by conveying the signal of “Don’t eat me” to CTL through the check point molecules (PD-L1, PD-1, etc.). Thus, the four chemokines might also indirectly enhance the tumor invasion and metastasis even when directly attacked by tumor cells. Since CCL5 was reported to rapidly induce cyclin D1, c-Myc, Ha-Ras through MARK-ERK, and Jak/STAT signaling, as well as glucose in-take and ATP production to stimulate tumor growth (Ding et al., 2016; D. ; Gao & Fish, 2018; Murooka, Rahbar, & Fish, 2009), the HNSCC cell lines with the high expression of the four chemokines were sensitive to the anti-tumor drug targeting MARK-ERK and RAS pathways.

The four chemokines were associated positively with Macrophages M1, activated CD4, and CD8 T cells, but negatively with Macrophages M0 (in TCGA HNSCC cohort) and tumor-associated monocyte/macrophages (GEO cohort), suggesting the lower expression of the four chemokines in the early tumor-associated monocyte/macrophages or macrophages M0, and the gradually elevated expression in differentiating TAMM. Moreover, the higher expression of the four chemokines were intensively detected in macrophage M1, instead of M2, also implicating their association with the polarization of macrophages, which was supported by the recent studies that found that the lack of CCL5 promoted the polarization in macrophages M2 (M. Li M et al., 2020), and CXCL9 and CXCL10 induced the polarization in macrophages M1 (Kohler et al., 2019). The scRNA-seq revealed that the mRNA peaks of the four chemokines in TAM were distributed in different subpopulations of TAMM, suggesting that the strong and exclusive co-regulation of the four chemokines disclosed by the bulk RNA-seq were inconsistent with the single cell level, which required further investigation.

In the immunotherapy cohort receiving anti-PD-L1/PD-1, the mRNA levels of the four chemokines were elevated significantly in CR group, verifying their positive roles in immune responses. It also encouraged us to establish a criterion predicting the patients’ responses to ICB therapy. In the premise of the substantial immune capacity of resisting tumors, ICB therapy facilitated T cells to eliminate tumor cells by blocking immune evasion, namely, the validity of ICB therapy depended on the patients’ immunity. Considering the crucial roles of the four chemokines (CXCL9, CXCL10, CXCL11, and CCL5) and the three pivotal immune cells (Macrophages M1, CD 4, and CD8 T cells) in tumor immunity, we established a criterion predicting the validity of ICB therapy through neural network, in which the AUC in training and test sets achieved 100% and 74%, respectively.

We have to acknowledge the shortcomings in this study. First, although they have been verified in other tumors, work is still required to verify the roles of the four chemokines in HNSCC found by the bioinformatical analyses. Second, the exploration on the mechanisms resulting in the immune evasion in the high immune infiltration subtype was insufficient. Third, the predicting and verifying cohorts were not HNSCC cohorts, which might weaken the prediction accuracy in HNSCC cohorts because of the variations among tumors. Fourth, despite the impressive promotion in tumor immunity, the roles of the four chemokines-CXCL9, CXCL10, CXCL11, and CCL5-in tumorigenesis and progression were still in debates.

CONCLUSION

In summary, we indeed established the core roles of the four chemokines-CXCL9, CXCL10, CXCL11, and CLL5-in HNSCC immunity by combining TAMM differentiation and HNSCC TME. From the perspective of the four chemokines associating TAMM with HNSCC immunity, we found the limitation of treating TAM as a static entity and the potential values of the four chemokines in tumor immunity. Considering the few studies on HNSCC immunity, our present study provided bioinformatics support for future explorations.

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

JX and CL designed the study. CL and HL wrote the manuscript. JX and XL revised and polished the manuscript. YH, HL, and XL performed the statistical analysis of the data, and generated the figures and tables. NL, HB, CG, TX, and LZ provided advice for the analysis and the figures. All authors reviewed and approved the final manuscript.

FUNDING

This study is supported by the National Natural Science Foundation of China (81771055).

ACKNOWLEDGMENTS

The authors appreciate TCGA database, GEO database, GDSC database, and IMvigor210 database for providing the original study data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.820065/full#supplementary-material

REFERENCES

Aldinucci, D., Borghese, C., and Casagrande, N. (2020). The CCL5/CCR5 Axis in Cancer Progression. *Cancers* 12 (7), 1765. doi:10.3390/cancers12071765

Bao, X., Shi, R., Zhao, T., Wang, Y., Anastasov, N., Rosemann, M., et al. (2021). Integrated Analysis of Single-Cell RNA-Seq and Bulk RNA-Seq Unravels Tumour Heterogeneity Plus M2-like Tumour-Associated Macrophage Infiltration and Aggressiveness in TNBC. *Cancer Immunol. Immunother.* 70 (1), 189–202. doi:10.1007/s00262-020-02669-7

Büttner, R., Longshore, J. W., López-Ríos, F., Merkelbach-Bruse, S., Normanno, N., Rouleau, E., et al. (2019). Implementing TMB Measurement in Clinical Practice: Considerations on Assay Requirements. *ESMO Open* 4 (1), e000442. doi:10.1136/esmoopen-2018-000442

Cao, B., Wang, Q., Zhang, H., Zhu, G., and Lang, J. (2018). Two Immune-Enhanced Molecular Subtypes Differ in Inflammation, Checkpoint Signaling and Outcome of Advanced Head and Neck Squamous Cell Carcinoma. *Oncoimmunology* 7 (2), e1392427. doi:10.1080/2162402x.2017.1392427

Cassetta, L., Fragkogianni, S., Sims, A. H., Swierczak, A., Forrester, L. M., Zhang, H., et al. (2019). Human Tumor-Associated Macrophage and Monocyte Transcriptional Landscapes Reveal Cancer-specific Reprogramming, Biomarkers, and Therapeutic Targets. *Cancer Cell* 35 (4), 588–602. e510. doi:10.1016/j.ccell.2019.02.009

Chan, T. A., Wolchok, J. D., and Snyder, A. (2015). Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *N. Engl. J. Med.* 373 (20), 1984. doi:10.1056/NEJMc1508163

Clarke, E., Eriksen, J. G., and Barrett, S. (2021). The Effects of PD-1/pd-L1 Checkpoint Inhibitors on Recurrent/metastatic Head and Neck Squamous Cell Carcinoma: a Critical Review of the Literature and Meta-Analysis. *Acta Oncologica* 60 (11), 1534–1542. doi:10.1080/0284186x.2021.1964699

Coifman, R. R., Lafon, S., Lee, A. B., Maggioni, M., Nadler, B., Warner, F., et al. (2005). Geometric Diffusions as a Tool for Harmonic Analysis and Structure Definition of Data: Diffusion Maps. *Proc. Natl. Acad. Sci. U.S.A.* 102 (21), 7426–7431. doi:10.1073/pnas.0500334102

Curran, M. A., Montalvo, W., Yagita, H., and Allison, J. P. (2010). PD-1 and CTLA-4 Combination Blockade Expands Infiltrating T Cells and Reduces Regulatory T and Myeloid Cells within B16 Melanoma Tumors. *Proc. Natl. Acad. Sci. U.S.A.* 107 (9), 4275–4280. doi:10.1073/pnas.0915174107

Dangaj, D., Bruand, M., Grimm, A. J., Ronet, C., Barras, D., Duttagupta, P. A., et al. (2019). Cooperation between Constitutive and Inducible Chemokines Enables T Cell Engraftment and Immune Attack in Solid Tumors. *Cancer Cell* 35 (6), 885–900. doi:10.1016/j.ccell.2019.05.004

Davoli, T., Uno, H., Wooten, E. C., and Elledge, S. J. (2017). Tumor Aneuploidy Correlates with Markers of Immune Evasion and with Reduced Response to Immunotherapy. *Science* 355 (6322), 8399. doi:10.1126/science.aaf8399

Ding, H., Zhao, L., Dai, S., Li, L., Wang, F., and Shan, B. (2016). CCL5 Secreted by Tumor Associated Macrophages May Be a New Target in Treatment of Gastric Cancer. *Biomed. Pharmacother.* 77, 142–149. doi:10.1016/j.biopha.2015.12.004

Efremova, M., Vento-Tormo, M., Teichmann, S. A., and Vento-Tormo, R. (2020). CellPhoneDB: Inferring Cell-Cell Communication from Combined Expression of Multi-Subunit Ligand-Receptor Complexes. *Nat. Protoc.* 15 (4), 1484–1506. doi:10.1038/s41596-020-0292-x

Estko, M., Baumgartner, S., Urech, K., Kunz, M., Regueiro, U., Heusser, P., et al. (2015). Tumour Cell Derived Effects on Monocyte/macrophage Polarization and Function and Modulatory Potential of Viscum Album Lipophilic Extract *In Vitro*. *BMC Complement. Altern. Med.* 15, 130. doi:10.1186/s12906-015-0650-3

Feng, B., Shen, Y., Pastor Hostench, X., Bieg, M., Plath, M., Ishaque, N., et al. (2020). Integrative Analysis of Multi-Omics Data Identified EGFR and PTGS2 as Key Nodes in a Gene Regulatory Network Related to Immune Phenotypes in Head and Neck Cancer. *Clin. Cancer Res.* 26 (14), 3616–3628. doi:10.1158/1078-0432.Ccr-19-3997

Franklin, R. A., Liao, W., Sarkar, A., Kim, M. V., Bivona, M. R., Liu, K., et al. (2014). The Cellular and Molecular Origin of Tumor-Associated Macrophages. *Science* 344 (6186), 921–925. doi:10.1126/science.1252510

Gao, D., and Fish, E. N. (2018). Chemokines in Breast Cancer: Regulating Metabolism. *Cytokine* 109, 57–64. doi:10.1016/j.cyto.2018.02.010

Gao, L., Wang, F.-q., Li, H.-m., Yang, J.-g., Ren, J.-G., He, K.-f., et al. (2016). CCL2/EGF Positive Feedback Loop between Cancer Cells and Macrophages Promotes Cell Migration and Invasion in Head and Neck Squamous Cell Carcinoma. *Oncotarget* 7 (52), 87037–87051. doi:10.18632/oncotarget.13523

García Campelo, M. R., Arriola, E., Campos Balea, B., López-Brea, M., Fuentes-Pradera, J., de Castro Carpeno, J., et al. (2021). PD-L1 Inhibitors as Monotherapy for the First-Line Treatment of Non-small-cell Lung Cancer in PD-L1 Positive Patients: A Safety Data Network Meta-Analysis. *Jcm* 10 (19), 4583. doi:10.3390/jcm10194583

Goswami, K. K., Ghosh, T., Ghosh, S., Sarkar, M., Bose, A., and Baral, R. (2017). Tumor Promoting Role of Anti-tumor Macrophages in Tumor Microenvironment. *Cell Immunol.* 316, 1–10. doi:10.1016/j.cellimm.2017.04.005

Guilliams, M., Ginhoux, F., Jakubzick, C., Naik, S. H., Onai, N., Schraml, B. U., et al. (2014). Dendritic Cells, Monocytes and Macrophages: a Unified Nomenclature Based on Ontogeny. *Nat. Rev. Immunol.* 14 (8), 571–578. doi:10.1038/nri3712

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., et al. (2021). Integrated Analysis of Multimodal Single-Cell Data. *Cell* 184 (13), 3573–3587. e3529. doi:10.1016/j.cell.2021.04.048

He, Y., Jiang, Z., Chen, C., and Wang, X. (2018). Classification of Triple-Negative Breast Cancers Based on Immunogenomic Profiling. *J. Exp. Clin. Cancer Res.* 37 (1), 327. doi:10.1186/s13046-018-1002-1

Hu, Y., He, M.-Y., Zhu, L.-F., Yang, C.-C., Zhou, M.-L., Wang, Q., et al. (2016). Tumor-associated Macrophages Correlate with the Clinicopathological Features and Poor Outcomes via Inducing Epithelial to Mesenchymal Transition in Oral Squamous Cell Carcinoma. *J. Exp. Clin. Cancer Res.* 35, 12. doi:10.1186/s13046-015-0281-z

Karnoub, A. E., Dash, A. B., Vo, A. P., Sullivan, A., Brooks, M. W., Bell, G. W., et al. (2007). Mesenchymal Stem Cells within Tumour Stroma Promote Breast Cancer Metastasis. *Nature* 449 (7162), 557–563. doi:10.1038/nature06188

Kim, M. H., Kim, J.-H., Lee, J. M., Choi, J. W., Jung, D., Cho, H., et al. (2020). Molecular Subtypes of Oropharyngeal Cancer Show Distinct Immune Microenvironment Related with Immune Checkpoint Blockade Response. *Br. J. Cancer* 122 (11), 1649–1660. doi:10.1038/s41416-020-0796-8

Kimura, Y. N., Watari, K., Fotovati, A., Hosoi, F., Yasumoto, K., Izumi, H., et al. (2007). Inflammatory Stimuli from Macrophages and Cancer Cells Synergistically Promote Tumor Growth and Angiogenesis. *Cancer Sci.* 98 (12), 2009–2018. doi:10.1111/j.1349-7006.2007.00633.x

Kloepper, J., Riedemann, L., Amoozgar, Z., Seano, G., Susek, K., Yu, V., et al. (2016). Ang-2/VEGF Bispecific Antibody Reprograms Macrophages and Resident Microglia to Anti-tumor Phenotype and Prolongs Glioblastoma Survival. *Proc. Natl. Acad. Sci. U.S.A.* 113 (16), 4476–4481. doi:10.1073/pnas.1525360113

Kohler, J. B., Cervilha, D. A. d. B., Riani Moreira, A., Santana, F. R., Farias, T. M., Alonso Vale, M. I. C., et al. (2019). Microenvironmental Stimuli Induce Different Macrophage Polarizations in Experimental Models of Emphysema. *Biol. Open* 8 (4), 808. doi:10.1242/bio.040808

Korniejewska, A., McKnight, A. J., Johnson, Z., Watson, M. L., and Ward, S. G. (2011). Expression and Agonist Responsiveness of CXCR3 Variants in Human T Lymphocytes. *Immunology* 132 (4), 503–515. doi:10.1111/j.1365-2567.2010.03384.x

Lee, C., Lee, J., Choi, S. A., Kim, S.-K., Wang, K.-C., Park, S.-H., et al. (2018). M1 Macrophage Recruitment Correlates with Worse Outcome in SHH Medulloblastomas. *BMC Cancer* 18 (1), 535. doi:10.1186/s12885-018-4457-8

Li, B., Ren, M., Zhou, X., Han, Q., and Cheng, L. (2020). Targeting Tumor-Associated Macrophages in Head and Neck Squamous Cell Carcinoma. *Oral Oncol.* 106, 104723. doi:10.1016/j.oraloncology.2020.104723

Li, M., Sun, X., Zhao, J., Xia, L., Li, J., Xu, M., et al. (2020). CCL5 Deficiency Promotes Liver Repair by Improving Inflammation Resolution and Liver Regeneration through M2 Macrophage Polarization. *Cell Mol Immunol* 17 (7), 753–764. doi:10.1038/s41423-019-0279-0

Movahedi, K., Laoui, D., Gysemans, C., Baeten, M., Stangé, G., Van den Bossche, J., et al. (2010). Different Tumor Microenvironments Contain Functionally Distinct Subsets of Macrophages Derived from Ly6C(high) Monocytes. *Cancer Res.* 70 (14), 5728–5739. doi:10.1158/0008-5472.Can-09-4672

Murooka, T. T., Rahbar, R., and Fish, E. N. (2009). CCL5 Promotes Proliferation of MCF-7 Cells through mTOR-dependent mRNA Translation. *Biochem. Biophysical Res. Commun.* 387 (2), 381–386. doi:10.1016/j.bbrc.2009.07.035

Packer, J. S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H., et al. (2019). A Lineage-Resolved Molecular Atlas of C Elegans Embryogenesis at Single-Cell Resolution. *Science* 365 (6459). doi:10.1126/science.aax1971

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., et al. (2014). Single-cell RNA-Seq Highlights Intratumoral Heterogeneity in Primary Glioblastoma. *Science* 344 (6190), 1396–1401. doi:10.1126/science.1254257

Romero, J. M., Grünwald, B., Jang, G.-H., Bavi, P. P., Jhaveri, A., Masoomian, M., et al. (2020). A Four-Chemokine Signature Is Associated with a T-Cell-Inflamed Phenotype in Primary and Metastatic Pancreatic Cancer. *Clin. Cancer Res.* 26 (8), 1997–2010. doi:10.1158/1078-0432.Ccr-19-2803

Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., and Hacohen, N. (2015). Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell* 160 (1-2), 48–61. doi:10.1016/j.cell.2014.12.033

Saloura, V., Izumchenko, E., Zuo, Z., Bao, R., Korzinkin, M., Ozerov, I., et al. (2019). Immune Profiles in Primary Squamous Cell Carcinoma of the Head and Neck. *Oral Oncol.* 96, 77–88. doi:10.1016/j.oraloncology.2019.06.032

Sanmamed, M. F., and Chen, L. (2018). A Paradigm Shift in Cancer Immunotherapy: From Enhancement to Normalization. *Cell* 175 (2), 313–326. doi:10.1016/j.cell.2018.09.035

Seiwert, T. Y., Burtness, B., Mehra, R., Weiss, J., Berger, R., Eder, J. P., et al. (2016). Safety and Clinical Activity of Pembrolizumab for Treatment of Recurrent or Metastatic Squamous Cell Carcinoma of the Head and Neck (KEYNOTE-012): an Open-Label, Multicentre, Phase 1b Trial. *Lancet Oncol.* 17 (7), 956–965. doi:10.1016/s1470-2045(16)30066-3

Seminerio, I., Kindt, N., Descamps, G., Bellier, J., Lechien, J. R., Mat, Q., et al. (2018). High Infiltration of CD68+ Macrophages Is Associated with Poor Prognoses of Head and Neck Squamous Cell Carcinoma Patients and Is Influenced by Human Papillomavirus. *Oncotarget* 9 (13), 11046–11059. doi:10.18632/oncotarget.24306

Shi, C., and Pamer, E. G. (2011). Monocyte Recruitment during Infection and Inflammation. *Nat. Rev. Immunol.* 11 (11), 762–774. doi:10.1038/nri3070

Singhal, S., Stadanlick, J., Annunziata, M. J., Rao, A. S., Bhojnagarwala, P. S., O'Brien, S., et al. (2019). Human Tumor-Associated Monocytes/macrophages and Their Regulation of T Cell Responses in Early-Stage Lung Cancer. *Sci. Transl. Med.* 11 (479), 1500. doi:10.1126/scitranslmed.aat1500

Spencer, K. R., Wang, J., Silk, A. W., Ganesan, S., Kaufman, H. L., and Mehnert, J. M. (2016). Biomarkers for Immunotherapy: Current Developments and Challenges. *Am. Soc. Clin. Oncol. Educ. Book* 35, e493–e503. doi:10.1200/edbk_160766

Tokunaga, R., Zhang, W., Naseem, M., Puccini, A., Berger, M. D., Soni, S., et al. (2018). CXCL9, CXCL10, CXCL11/CXCR3 axis for Immune Activation - A Target for Novel Cancer Therapy. *Cancer Treat. Rev.* 63, 40–47. doi:10.1016/j.ctrv.2017.11.007

Topalian, S. L., Hodi, F. S., Brahmer, J. R., Gettinger, S. N., Smith, D. C., McDermott, D. F., et al. (2012). Safety, Activity, and Immune Correlates of Anti-PD-1 Antibody in Cancer. *N. Engl. J. Med.* 366 (26), 2443–2454. doi:10.1056/NEJMoa1200690

Tummers, B., and Green, D. R. (2017). Caspase-8: Regulating Life and Death. *Immunol. Rev.* 277 (1), 76–89. doi:10.1111/imr.12541

Turan, T., Kongpachith, S., Halliwill, K., Roelands, J., Hendrickx, W., Marincola, F. M., et al. (2021). A Balance Score between Immune Stimulatory and Suppressive Microenvironments Identifies Mediators of Tumour Immunity and Predicts Pan-Cancer Survival. *Br. J. Cancer* 124 (4), 760–769. doi:10.1038/s41416-020-01145-4

Walter, V., Yin, X., Wilkerson, M. D., Cabanski, C. R., Zhao, N., Du, Y., et al. (2013). Molecular Subtypes in Head and Neck Cancer Exhibit Distinct Patterns of Chromosomal Gain and Loss of Canonical Cancer Genes. *PLoS One* 8 (2), e56823. doi:10.1371/journal.pone.0056823

Wang, S.-W., Wu, H.-H., Liu, S.-C., Wang, P.-C., Ou, W.-C., Chou, W.-Y., et al. (2012). CCL5 and CCR5 Interaction Promotes Cell Motility in Human Osteosarcoma. *PLoS One* 7 (4), e35101. doi:10.1371/journal.pone.0035101

Watters, J. J., Schartner, J. M., and Badie, B. (2005). Microglia Function in Brain Tumors. *J. Neurosci. Res.* 81 (3), 447–455. doi:10.1002/jnr.20485

Xiao, M., Zhang, J., Chen, W., and Chen, W. (2018). M1-like Tumor-Associated Macrophages Activated by Exosome-Transferred THBS1 Promote Malignant Migration in Oral Squamous Cell Carcinoma. *J. Exp. Clin. Cancer Res.* 37 (1), 143. doi:10.1186/s13046-018-0815-2

Yang, C., Zheng, W., and Du, W. (2016). CXCR3A Contributes to the Invasion and Metastasis of Gastric Cancer Cells. *Oncol. Rep.* 36 (3), 1686–1692. doi:10.3892/or.2016.4953

Yi, L., Wu, G., Guo, L., Zou, X., and Huang, P. (2020). Comprehensive Analysis of the PD-L1 and Immune Infiltrates of m6A RNA Methylation Regulators in Head and Neck Squamous Cell Carcinoma. *Mol. Ther. - Nucleic Acids* 21, 299–314. doi:10.1016/j.omtn.2020.06.001

Zhang, H., Luo, Y.-B., Wu, W., Zhang, L., Wang, Z., Dai, Z., et al. (2021). The Molecular Feature of Macrophages in Tumor Immune Microenvironment of Glioma Patients. *Comput. Struct. Biotechnol. J.* 19, 4603–4618. doi:10.1016/j.csbj.2021.08.019

Zhang, X., Shi, M., Chen, T., and Zhang, B. (2020). Characterization of the Immune Cell Infiltration Landscape in Head and Neck Squamous Cell Carcinoma to Aid Immunotherapy. *Mol. Ther. - Nucleic Acids* 22, 298–309. doi:10.1016/j.omtn.2020.08.030

Zheng, Y., Bao, J., Zhao, Q., Zhou, T., and Sun, X. (2018). A Spatio-Temporal Model of Macrophage-Mediated Drug Resistance in Glioma Immunotherapy. *Mol. Cancer Ther.* 17 (4), 814–824. doi:10.1158/1535-7163.Mct-17-0634

Ziegler-Heitbrock, L., Ancuta, P., Crowe, S., Dalod, M., Grau, V., Hart, D. N., et al. (2010). Nomenclature of Monocytes and Dendritic Cells in Blood. *Blood* 116 (16), e74–e80. doi:10.1182/blood-2010-02-258558

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership