**Abstract**

**Purpose:** The purpose of this study was to construct a model based on the prognostic features associated with epithelial-mesenchymal transition (EMT) to generate new ideas for exploring the mechanism and treatment of hepatocellular carcinoma (HCC) metastasis and invasion.

**Methods:** Consistent clustering analysis was used to identify EMT-associated genes and molecular subtypes. Differentially expressed genes (DEGs) between molecular subtypes were calculated using the limma package and then subjected to functional enrichment analysis. The immune cell scores were compared amongst the molecular subtypes using the ESTIMATE, MCP counter, and GSCA of R package, respectively. A multi-gene prognostic model was constructed using lasso regression, and immunotherapy effects of the model were analyzed using the Imvigor210 cohort. In addition, to validate gene expression, Immunohistochemical analysis was performed on a real hepatocellular carcinoma cohort.

**Results:** Based on 59 EMT-associated genes, 365 LIHC samples were classified into two subtypes, with the C1 subtype having a poorer prognosis and having higher immune scores than the C2 subtype, with more associated with the tumor progression pathways. Based on 1130 DEGs between subtypes, a 4-gene signature was constructed that had strong robustness and performed stable predictive efficacy in different platform datasets (HCCDB18 and GSE14520). Comparing with other existing models, this model had better performance. The immunotherapy cohort Imvigor210 was used to validate the potential of gene signature to predict immunotherapy response, which revealed that the FTCD, PON1, and TMEM45A genes were all significantly overexpressed in cancerous tissues, whereas the G6PD gene expression was significantly low.

**Conclusion:** The 4-genes signature based on the EMT-associated genes provided valuable information for future research of the pathogenesis and clinical management of HCC.

Keywords: ???

1. **Introduction**

Liver cancer is the sixth most common type of cancer in the world and the third leading cause of cancer-related deaths, with high morbidity and mortality as well as an extremely poor prognosis [1]. HCC is a predominant histologic form of liver cancer, accounting for 85-90% of all primary liver cancers and causing 700,000 deaths worldwide each year, which is more prevalent and fatal in developing countries [2-4]. For individuals with advanced HCC, sorafenib has been the sole systemic therapeutic choice. Combination therapy with atezolizumab and bevacizumab became a new frontline standard of care for unresectable or metastatic HCC in 2020 [5]. Despite recent advancements in treatment, the overall 5-year survival rate is now less than 12%, mainly associated with high recurrence rates, and associated intra- or extrahepatic metastases. The majorities of patients with HCC are advanced at diagnosis and have limited clinical benefit from treatment [6-9]. Since HCC has a rather poor prognosis and is highly resistant to most anticancer therapies, research has been undertaken to unravel the complex molecular mechanisms of hepatocarcinogenesis and progression, such as EMT, tumor-stromal interactions, tumor microenvironment, tumor stem cells, and senescence bypass [10]. A better understanding of these mechanisms may have implications for the development of novel and more effective therapeutic and prognostic strategies that are urgently needed.

EMT is an important biological process that occurs throughout embryonic development, cell differentiation and reprogramming, and cancer progression [11, 12]. Several studies have reported that EMT confers tumor stem cell-like features that enable treatment resistance and tumor recurrence [13]. It is considered one of the main mechanisms determining cancer cell invasion and metastasis [14]. Many studies have also found that EMT is associated with the invasion and progression of various malignancies, including HCC [15, 16]. Similar to other tumors, EMT in HCC cells appears to be mediated by aberrant activation of the Wnt/β-catenin signaling pathway [17-20], which promotes hypoxia-induced EMT in HCC [21]. EMT has been found to aid cell proliferation, invasion, and metastasis during HCC progression, as well as contribute to HCC metastasis and chemotherapy resistance. EMT leads to HCC metastasis and poor patient prognosis [22-24]. EMT also positively correlates with sorafenib, cisplatin, and Adriamycin resistance [25-27], and has a significant negative impact on the survival of HCC patients by inducing metastasis and resistance, which ultimately leads to poor prognosis. Sorafenib inhibits HCC cells migration by suppressing EMT, which is one of the potential mechanisms for the antitumor effect of sorafenib in HCC [28]. Despite extensive research into the mechanism of action of EMT in HCC, the prognostic value and biological relevance of EMT-associated genes remain unknown. Therefore, studying the molecular subtyping of HCC associated with EMT and its prognostic relevance is critical for identifying therapeutic targets and improving the prognosis of HCC patients.

**IV.Discussion**

Patients with HCC typically have no clinically significant symptoms in the early stages, and it remains a major public health challenge worldwide due to its high morbidity and mortality [33]. Given the enormous heterogeneity of HCC, the identification of new prognostic markers and the construction of more accurate prognostic models are urgently needed.

In this study, 365 LIHC samples from The Cancer Genome Atlas (TCGA) were classified into two molecular subtypes based on 59 EMT-associated genes, with distinct clinical features and prognostic outcomes. In general, the C1 group with a poorer prognosis had a higher proportion of deaths, a higher T-stage, a higher degree of differentiation, more advanced staging, and higher immune scores. Based on this, a prognostic evaluation model was constructed, which can distinguish the different molecular subtypes as well as can evaluate the prognosis of patients with HCC better than previously developed methods.

There has been an increasing number of studies on tumor prognostic models, but there are no clear findings on predicting prognosis based on EMT-associated markers in HCC. In this study, a novel 4-gene marker (including PON1, FTCD, G6PD, and TMEM45A) with strong robustness was developed for HCC prognosis prediction based on EMT-associated molecules and validated in two other independent cohorts. Paraoxonase-1 (PON1), a Ca2+-dependent high-density lipoprotein (HDL)-associated endostatin, is the first member of the paraoxonase (PON) multigene family, which is involved in the antioxidant function of HDL with atheroprotective effects and is associated with the pathogenesis of many diseases, including cardiovascular disease and cancer [34-36]. Several studies have shown that PON1 activity is associated with the progression of many cancers [37, 38]. PON1 gene polymorphism is associated with breast cancer susceptibility [39], while PON1 concentration is positively correlated with the degree of bone destruction in multiple myeloma [40]. PON1 activity is increased in the serum of patients with colorectal cancer and is also higher in colon cancer tissues [41]. Serum PON1 concentration is considerably decreased after radiotherapy and may be utilized as a marker of radiotherapy efficacy [42, 43]...PON1 has also been extensively researched in the field of HCC [44-48], and serum PON1 is now used as a biomarker to evaluate micro-vascular infiltration in HCC [49, 50]. In addition to its role in metabolism, the FTCD (formiminotransferase cyclodeaminase) catalyzes the degradation of histidine during folate metabolism and is also associated with the Golgi complex [51]. The FTCD gene is a potential tumor suppressor gene for HCC [52], which is significantly downregulated in HCC tumor tissues and can be used as a diagnostic biomarker to distinguish early-stage HCC from benign tumors [53]. The combined expression of arginase 1 + FTCD + MOC 31 contributes to the diagnosis of most hepatocellular and metastatic cancers [54]. Furthermore, FTCD was found to correlate with methotrexate chemotherapy drug sensitivity [55]. Reduced NADPH produced by glucose-6-phosphate dehydrogenase (G6PD) is essential for the maintenance of intracellular redox homeostasis and reductive biosynthesis, and G6PD deficiency is one of the most common inherited enzyme deficiency disorders [56]. G6PD, the rate-limiting enzyme of the pentose phosphate pathway, is commonly activated in human malignancies to produce precursors for nucleotide and lipid synthesis, and abnormal activation of G6PD leads to the proliferation of a variety of cancer cells [57]. G6PD activity is increased in several cancer types, including esophageal, gastric, colorectal, bladder, breast, and lung cancers [58-60]. Increased levels of G6PD mRNA expression indicate poor clinical outcomes, such as increased drug resistance, tumor cell migration, or proliferation. Therefore, it is expected that G6PD will become a viable target for oncology therapy soon [61]. G6PD was found to be an important miR-122 target that may regulate glucose metabolism in HCC, and upregulation of G6PD, and was observed to be associated with higher tumor grade, increased tumor recurrence, and poorer patient survival [62]. TMEM45A belongs to the family of transmembrane proteins (TMEM), which is predicted to be components of various cell membranes, such as the mitochondrial membrane, the endoplasmic reticulum membranes, and Golgi membranes [63]. Under hypoxic conditions, chemotherapy resistance in human breast cancer cells and HCC cells is related, which also affects the proliferation and invasion of human ovarian cancer cells and human glioma cells [64-67]. By blocking the TGF-signaling pathway in human colorectal cancer cells, TMEM45A gene knockdown proved efficient in inhibiting multidrug resistance and suppressing EMT [68]. These studies suggest that TMEM45A may be a potential biomarker. The 4-gene marker identified in this study is involved in a variety of tumorigenicity processes and is closely associated with HCC tumor cell proliferation, metastasis, or invasion. This 4-gene signature could be a powerful biomarker for predicting the prognosis of HCC.

Notably, GSEA showed highly enhanced tumor characteristics as well as various metabolic features. The findings revealed that a large number of tumor-related pathways were significantly overexpressed in the poor prognosis subtype C1, implying that C1 subtype tumors are more aggressive. These findings are also consistent with the clinical features of C1 subtype tumors such as late-stage, high differentiation degree, and high mortality. On the other hand, the expression levels of metabolism-related pathways were higher in C2 subtypes with better prognosis. Most of these metabolic pathways were related to physiological hepatocyte metabolic functions, such as fatty acids, PPAR signaling pathway, and drug metabolic processes, indicating a more intact hepatocyte function, thus contributing to the clinical outcome.

Furthermore, tumor-related pathways increased with increasing Risk Scores, while metabolism-related pathways decreased with increasing Risk Scores. This also indicated that Risk Scores can predict the prognosis of HCC and help us better understand the underlying molecular mechanisms of hepatocellular carcinogenesis and progression.

Three published gene signatures for HCC were compared to demonstrate the superiority of this model. In a previous study, 149 pairs of HCC specimens were obtained from GEO, 98 DEGs were screened between HCC and normal hepatic tissue, established and a 4-gene subset of prognostic gene expression signature for HCC (SPINK1, TXNRD1, LCAT, and PZP) was validated. The results indicated that the expression panel of these four genes was strongly correlated with methylation status [30]. A 6-gene signature [31] identified two prognostic molecular subtypes of HCC with different expression profiles and clinical outcomes and established a prognostic evaluation model that distinguished different subtypes of HCC and also provided a good evaluation of patient prognosis. Another 6-gene signature [32] established a new 6-gene marker (including CSE1L, CSTB, MTHFR, DAGLA, MMP10, and GYS2) that classified the HCC patients into high-and low-risk groups with significant differences in survival rates. Their rates of survival differed significantly. The ROC analysis of the four models showed that the 5-year AUC values of the previously developed 4-gene signature [30], 6-gene signature [31], and 6-gene signature [32] were lower than this newly developed model in this study, indicating that this model is more reasonable and effective with a reasonable number of genes. Moreover, the C-index values of our Risk Score model were higher than those of the other three models, demonstrating the new model’s superior performance.

PD-L1 expression, tumor mutation burden (TMB), and DNA mismatch repair defects have all been identified as genetic markers associated with cancer immunotherapy response [69-71].In 2020, the NCCN guidelines emphasized the use of atezolizumab and bevacizumab combination therapy [5], although the current immunotherapy for HCC is limited in terms of effective predictive markers [72, 73]. The prediction of immunotherapy efficacy by the 4-genes model was explored by the immunotherapy dataset (Imvigor210), which revealed that patients in the CR group had a significantly lower Risk Score than the PD group. The higher Risk Score values were associated with poorer survival. Moreover, the proportion of samples with immunotherapy response (CR+PR) was smaller in the high-risk group than in the low-risk group (18% vs. 26%), which suggested that HCC patients from the high-Risk Score group may not respond effectively to immunotherapy; this needs to be investigated in future clinical trials.