Background: Hepatocellular carcinoma (HCC) is one of the most common malignant tumors worldwide, and its prognosis remains unsatisfactory. The identification of new and effective markers is helpful for better predicting the prognosis of patients with HCC and for conducting individualized management. The oncogene Aurora kinase A (AURKA) is involved in a variety of tumors; however, its role in liver cancer is poorly understood. The aim of this study was to establish AURKA-related gene signatures for predicting the prognosis of patients with HCC.

Methods: We first analyzed the expression of AURKA in liver cancer and its prognostic significance in different data sets. Subsequently, we selected genes with prognostic value related to AURKA and constructed a gene signature based on them. The predictive ability of the gene signature was tested using the HCC cohort development and verification data sets. A nomogram was constructed by integrating the risk score and clinicopathological characteristics. Finally, the influence of the gene signature on the immune microenvironment in HCC was comprehensively analyzed.

Results: We found that AURKA was highly expressed in HCC, and it exhibited prognostic value. We selected eight AURKA-related genes with prognostic value through the protein-protein interaction network and successfully constructed a gene signature. The nine-gene signature could effectively stratify the risk of patients with HCC and demonstrated a good ability in predicting survival. The nomogram showed good discrimination and consistency of risk scores. In addition, the high-risk group showed a higher percentage of immune cell infiltration (i.e., macrophages, myeloid dendritic cells, neutrophils, and CD4+T cells). Moreover, the immune checkpoints SIGLEC15, TIGIT, CD274, HAVCR2, and PDCD1LG2 were also higher in the high-risk group versus the low-risk group.

Conclusions: This gene signature may be useful prognostic markers and therapeutic targets in patients with HCC.

Keywords: AURKA, gene signature, hepatocellular carcinoma, prognosis, immune infiltration, nomogram
INTRODUCTION

According to global data, hepatocellular carcinoma (HCC) is the most common primary liver tumor and the third most common risk factor for cancer-related deaths worldwide (Sung et al., 2021). In 2020, there were approximately 906,000 newly diagnosed patients with liver cancer worldwide and approximately 830,000 liver cancer-related deaths (Sung et al., 2021). The main causes of liver cancer include hepatitis virus infection, smoking, alcoholic cirrhosis, chemical drugs, and aflatoxin infection (Forner et al., 2018). Approximately 400,000 people die annually in China, accounting for >50% of liver cancer-related deaths globally (Chen et al., 2016). Early-stage HCC is insidious and difficult to detect; consequently, a large number of patients already have advanced-stage disease at the time of diagnosis. At present, liver cancer is mainly treated by surgical resection, supplemented by other methods, such as ablation therapy, targeted therapy, and immunotherapy (Hartke et al., 2017). However, even with timely intervention, the recurrence and mortality rates remain high due to the high degree of malignancy in liver cancer, rapid disease progression, and poor prognosis (Li et al., 2015). There are numerous markers used for predicting the prognosis of patients with liver cancer; nevertheless, their effectiveness is currently limited. Therefore, there is an urgent need to identify more effective biomarkers for predicting the prognosis of patients with liver cancer.

Aurora kinase A (AURKA) is a member of the Aurora kinase family, which consists of AURKA A, B, and C (Carmena and Earnshaw, 2003). Human AURKA is located on chromosome 20q13 and encodes a protein of 403 amino acids. It mainly regulates mitotic spindle formation, stability, and chromosome segregation, and plays an important role in cell cycle regulation (Lindon et al., 2016). The abnormal expression of AURKA can lead to chromosomal abnormalities and instability of the cell genome, which is a risk factor for tumor formation (Wu et al., 2018). AURKA is abnormally expressed in a variety of tumors and regulates tumor proliferation, migration, invasion, and metastasis (Yan et al., 2016). In addition, it is involved in multiple signaling pathways, such as the TP53 pathway, Ras/mitogen-activated protein kinase (MAPK) pathway, and NFκB pathway (Katayama et al., 2004; Briassouli et al., 2007; Umstead et al., 2017). Previous studies have confirmed that AURKA is related to the prognosis of a variety of cancers (breast, colorectal, pancreatic, gastric, and head and neck) and may be a therapeutic target (Jeng et al., 2004; Reiter et al., 2006; Zhang et al., 2015).

Investigations showed that AURKA played an important role in liver cancer progression. Jeng et al. confirmed that AURKA was overexpressed frequently and correlated with high grade and high stage in HCC (Jeng et al., 2004). Lu et al. reported that AURKA mediated c-Myc’s oncogenic effects in HCC (Lu et al., 2015). Zhang et al. revealed AURKA promoted chemoresistance through targeting NF-kappaB/microRNA-21/PTEN signaling pathway in HCC (Zhang et al., 2014). Chen et al. suggested that AURKA promoted cancer metastasis through regulating epithelial-mesenchymal transition and cancer stem cell properties in HCC (Chen et al., 2017). However, the specific mechanism of AURKA in HCC is still not very clear and needs to be further explored.

Thus far, no study investigated the role of AURKA gene and AURKA-related prognostic genes in liver cancer. It is well established that the tumor immune microenvironment influences tumor progression (Locy et al., 2018). Currently, the immunological value of AURKA in liver cancer has not been reported. In this study, we first analyzed its clinical value in HCC and selected prognostic genes associated with AURKA. Furthermore, we developed an AURKA-related gene signature in HCC. Next, we constructed a nomogram by combining risk scores and clinical characteristics. Finally, we evaluated the relationship between the gene signature and tumor immunity in HCC.

METHODS

Identification of AURKA as a Differentially Expressed Gene (DEG)

The RNA-seq data of LIHC patients from The Cancer Genome Atlas (TCGA, http://gdc.cancer.gov/) database and three datasets, including GSE14323 (HCC, n = 55; normal, n = 60), GSE14520 (HCC, n = 225; normal, n = 220) and GSE25097 (HCC, n = 268; normal, n = 289), from Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) database was downloaded to analysis the AURKA expression in LIHC patients. Gene expression levels were normalized by Robust Multi-Array Average (RMA).

Differential gene expression analysis of mRNAs was performed based on TCGA database. Limma package of R software (R version 3.6.2) was used to conduct differential gene expression analysis. The adjusted p-value was analyzed to correct for false positive results in TCGA or GTEx. “Adjusted p < 0.05 and Log (Fold Change) > 1 or Log (Fold Change) < -1” were regarded as the thresholds of differential expression of mRNAs. The study flowchart was presented in Figure 1.

Protein-Protein Interaction Network and Gene Enrichment Analysis

STRING (version 11.0; http://string.embl.de/) is an open-access biological database that predicts comprehensive interactions of genes at the protein level from multiple organisms including Homo sapiens (Mering et al., 2003). To screen AURKA-related genes, we used STRING to explore AURKA-related genes as well as conduct PPI network analysis on AURKA-related genes. The protein–protein interactions (PPI) with medium confidence >0.4 were regarded as significant.

To further confirm the underlying function of potential targets, the data were analyzed by functional enrichment. Gene Ontology (GO) is a widely-used tool for annotating genes with functions, especially molecular function (MF), biological pathways (BP), and cellular components (CC). Kyoto Encyclopedia of Genes and Genomes (KEGG) Enrichment Analysis is a practical resource for analytical study of gene functions and associated high-level genome functional
information. ClusterProfiler package (version: 3.18.0) in R was employed to analyze the GO function of potential targets and enrich the KEGG pathway. \( p < 0.05 \) was set as the cut-off criterion.

**Identification of Prognostic Genes**

Gene Expression Profiling Interactive Analysis (GEPIA) is an interactive web application based on The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression databases (Tang et al., 2017). We used the Gene Expression Profiling Interactive Analysis to screen the prognostic value of AURKA and AURKA-related genes. The identification of prognostic genes was based on the following criteria: 1) significant differences in gene expression levels between normal liver samples and liver tumor samples and 2) significant association of the gene with both overall survival (OS) and disease-free survival (DFS) of patients with liver hepatocellular carcinoma (LIHC). \( p \)-values <0.05 denoted statistically significant differences.

**Development and Validation of the Gene Signature**

Primary screening of the LIHC data from TCGA database was performed for missing data. After deleting the samples with missing data, The available LIHC data from the TCGA database (340 samples) was divided into two subsets: a training set (\( n = 240 \)) and a validation set (\( n = 100 \)) randomly using the random function of Microsoft Excel program. The training set was used to train a predictive model and the test set was applied for validation. Another database International Cancer Genome Consortium (ICGC) (dcc.icgc.org) was also used as validation set.

Identified prognostic genes were submitted to the multivariate Cox regression model to calculate each prognostic gene's coefficient and risk scores. We used X-tile plot (version 3.6.1, http://www.tissuearray.org/rimmlab) to determine the optimum cutoff of AURKA-related gene signature risk score. X-tile which could calculate the best cut-point of sub-populations, is a software developing by team Rimm Laboratory from Yale University (Camp et al., 2004). A train set and two validation sets were divided into three sub-groups using the same best cut-off value. Risk score analysis, including risk score distribution, survival status, and gene expression heatmap among sub-groups were performed. Kaplan–Meier curves were analyzed using Kruskal-Wallis test and visualized through GraphPad Prism (version 8.0). Receiver operating characteristic (ROC) curve analysis was performed to assess the predictive value of AURKA-related gene signature. The clinical characteristics of three cohorts were shown in Table 1. The correlation among clinicopathological characteristics and risk groups were analyzed by the chi-square test. For sample sizes of less than 40 or theoretical frequencies (\( T \)) of less than 1, Fisher exact probability method was used.

**Distribution of Somatic Mutations**

To identify the somatic mutations of patients with LIHC, we downloaded single-nucleotide polymorphism (SNV) data and clinical follow-up information from TCGA database. The downloaded single-nucleotide polymorphism data were organized in the multiple alignment (MAF) format and visualized using the “maftools” package in R software. The horizontal histogram showed the genes with the highest frequency of mutation.
Immune Infiltration Analysis
To explore the associations between different subgroups and immune cells infiltration, we employed Tumor Immune Estimation Resource (TIMER), which is a useful resource for comprehensive analysis of tumor-infiltrating immune cells (Li et al., 2017). The infiltration of six type of immune cell, including B cell, Macrophage, Myeloid dendritic cell, Neutrophil, T cell CD4+ and T cell CD8+, were calculated. SIGLEC15, IDO1, CD274, HAVCR2, PDCD1, CTLA4, LAG3 and PDCD1LG2 were selected to be immune checkpoints and the expression values of these eight immune checkpoints among sub-groups were explored. Differences between the three groups were assessed using the Kruskal-Wallis test. p < 0.05 was considered statistically significant. All the above analysis methods and R package were implemented by R foundation for statistical computing (2020) version 4.0.3 and software packages ggplot2 and pheatmap.

Construction and Validation of a Prognostic Model
A nomogram model was constructed to predict the probability of survival at 3 and 5 years for liver cancer patients. Briefly, the prognostic value of the clinicopathological characteristics for OS was estimated through univariate and multivariate Cox regression analyses in both training set and validation set. The performance of the risk model was validated by internal validation and external validation. Internal validation was performed by bootstrap Cox proportional regression analysis based on 1,000 bootstrap samples. Validation set was conducted based on another HCC patients from the TCGA database. Those parameters with p-values <0.05 in both training set and validation set were identified as potential prognostic factors, which were were included in multivariate Cox regression model and visualized using R package “rms.” The Calibration curves were plotted to analyze the diagnostic performance of the nomogram. The ROC curve were conducted to determine the clinical value of the nomogram.

RESULTS

AURKA mRNA Level in HCC Samples
The results of the DEG analysis are shown in Figure 2. In TCGA database, a total of 421 HCC samples were selected for the DEG analysis. The analysis showed that the AURKA gene was upregulated (log2 fold change >1.5 and adjusted p-value <0.05), which indicated that the mRNA expression levels of AURKA differ significantly between normal liver tissue and liver cancer tissue. The mRNA expression levels of AURKA were also determined using three Gene Expression Omnibus data sets (GSE14323, n = 115; GSE14520, n = 445; and GSE25097, n = 557). The results also showed that AURKA was significantly highly expressed in HCC.
PPI Network Analysis

According to the predictive results of the Search Tool for the Retrieval of Interacting Genes (STRING) database, another ten genes were identified as AURKA-related genes with significant interaction, namely targeting protein for Xklp2 (TPX2), cyclin dependent kinase 1 (CDK1), polo like kinase 1 (PLK1), DLG associated protein 5 (DLGAP5), cell division cycle 20 (CDC20), baculoviral IAP repeat containing 5 (BIRC5), transforming acidic coiled-coil containing protein 3 (TACC3), centromere protein A (CENPA), cyclin B2 (CCNB2), and ubiquitin conjugating enzyme E2 C (UBE2C). The protein-protein interaction (PPI) network of AURKA and AURKA-related genes was constructed and visualized using the online STRING database (Figure 3A).

The Gene Ontology (GO) enrichment analysis was composed of three parts: GO biological process (GO-BP), GO cellular component (GO-CC), and GO molecular function (GO-MF). In GO-CC (Figure 3B), these genes were enriched in condensed nuclear chromosome, centromeric region, mitotic spindle, and spindle pole. In GO-BP (Figure 3C), AURKA and its related genes were significantly enriched in mitotic nuclear envelope disassembly, positive regulation of ubiquitin protein ligase activity, and anaphase-promoting complex-dependent catabolic process. In GO-MF (Figure 3D), genes were mainly enriched in protein serine/threonine kinase activity, protein heterodimerization activity, and protein kinase binding. For the Kyoto Encyclopedia of Genes and Genomes pathway analysis, nine pathways (Figure 3E) were observed, namely...
progesterone-mediated oocyte maturation, oocyte meiosis, cell cycle, TP53 signaling pathway, forkhead box O (FOXO) signaling pathway, ubiquitin-mediated proteolysis, cellular senescence, viral carcinogenesis, and human T-lymphotropic virus type I (HTLV−I) infection.

**Development and Validation of a Nine-Gene Signature**

The results of the identification of prognostic genes are shown in Figure 4. According to the screening strategy and criteria described above, two AURKA-related genes, namely CCNB2 (OS Log-rank p > 0.05) and UBE2C (OS Log-rank p > 0.05), were excluded. Eight other AURKA-related genes were significantly highly expressed in HCC tissue compared with normal liver tissue and correlated with prognosis.

The AURKA and the other eight AURKA-related genes were used to construct a risk model. The risk score was calculated through multivariate regression analyses. The cutoff value of the risk score was identified using the X-tile software. According to the results, the HCC samples were divided into three subgroups. The cutoff values of the
risk score were 0.17 and 0.95. HCC samples with a risk score < 0.17 and > 0.95 classified into the low- and high-risk groups, respectively. The remaining samples were assigned to the moderate-risk group.

The details of the risk model showed in Figure 5, which revealed that higher risk scores were associated with higher expression levels of the nine-gene signature. Furthermore, higher risk scores also indicated worse OS. These results were similar in both the training set and other two validation sets. Collectively, these results suggest that the risk model had potential value in predicting the prognosis of HCC. Furthermore, as the ROC curves showed, the area under the ROC curve (AUC) of the training set and two validation sets were higher than 0.5, which indicated that AURKA-related gene signature risk model had important value in predicting prognosis (Figures 5D–F). The clinicopathological characteristics among the risk groups in training set were shown in Table 2. Age, T stage, pathological TNM (pTNM) stage, grade, recurrence, and survival endpoint were significantly different between the three risk groups (p < 0.05).
We compared the expression levels of activating 3 (LAG3), and programmed cell death 1 ligand 2 T-lymphocyte associated protein 4 (CTLA4), lymphocyte (HAVCR2), programmed cell death 1 (PDCD1), cytotoxic domains (TIGIT), CD274, hepatitis A virus cellular receptor 2 lectin 15 (SIGLEC15), T cell immunoreceptor with Ig and ITIM immune-related genes were selected: sialic acid binding Ig like 15 (SIGLEC15), T cell immunoreceptor with Ig and ITIM domains (TIGIT), CD274, hepatitis A virus cellular receptor 2 (HAVCR2), programmed cell death 1 (PDCD1), cytotoxic T-lymphocyte associated protein 4 (CTLA4), lymphocyte activating 3 (LAG3), and programmed cell death 1 ligand 2 (PDCD1LG2). We compared the expression levels of immune-related genes among the three risk groups. The results are shown in Figure 8. Except for CTLA4, LAG3, and PDCD1, the expression levels of the other five immune-related genes differed significantly between the three groups. Overall, the high-risk group showed significantly higher levels of immune gene expression and immune cell infiltration compared with the other groups.

### Construction and Evaluation of the Nomogram Model

As shown in Table 3, we performed the univariate and multivariate analyses using the SPSS software (version 23.0; IBM Corporation, Armonk, NY, United States) to identify independent prognostic factors predicting OS in patients with HCC. In the training data set, the risk score, T stage, M stage, pTNM stage, and recurrence were identified as independent prognostic factors. The results of the testing and training data sets were similar. Overall, the risk score, T stage, and pTNM stage were identified as independent prognostic factors in both data sets.

Hence, we used age, T stage, N stage, M stage, pTNM stage, recurrence, and risk group as estimated factors in the construction of our model. A nomogram was constructed to estimate the probabilities for three- and 5-year survival. Calibration curves analysis showed that the new nomogram model had good predictive accuracy (Figure 9B). Furthermore, the new nomogram model achieved an area under the curve (AUC) of 0.78 at 5 years, which was better than that of a model without the gene signature (AUC = 0.73) or pTNM stage (AUC = 0.70) (Figure 9C).

### DISCUSSION

The prognosis of patients with HCC varies greatly. The 5-year survival rate after resection of early HCC can be as high as 70%, whereas that of patients with vascular invasion or advanced HCC is markedly lower (Mazzaferro et al., 2011; Grandhi et al., 2016). Therefore, the use of effective prognostic indicators can promptly identify high-risk patients and assist in implementing individualized treatment to improve the prognosis.

In the present study, we found that the AURKA gene was highly expressed in HCC and an independent prognostic risk factor. The GO analysis of AURKA and related genes indicated that these genes are involved in a variety of biological processes, including nuclear chromosome condensation, centromeric regions, mitotic spindle and spindle poles, mitotic nuclear envelope disassembly, and ubiquitin protein ligase activity. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis showed that AURKA and related genes regulate the progression of HCC through multiple pathways, such as the TP53 signaling pathway and FOXO signaling pathway. Numerous studies have demonstrated that the TP53 signaling pathway is involved in the development of a variety of tumors and plays a regulatory role in tumor immunity (Stegh, 2012; Blagih et al., 2020; Muñoz-Fontela et al., 2016). The role of the FOXO

![Figure 6: Somatic Mutation Results](image_url)

Figure 6 illustrates the somatic landscape of the three risk subgroups. Information on the mutation status of each gene in each sample was shown in the waterfall plot, where different colors with specific annotations at the bottom indicated the various types of mutation. The barplot above the legend exhibited the number of mutations. The results showed that CTNNB1 was the most commonly mutated gene in the low-risk group. Tumor protein p53 (TP53) was the most frequently mutated gene in both the moderate- and high-risk groups. Hence, we further grouped the HCC samples into two groups based on the TP53 mutation status. A total of 101 and 241 HCC samples were assigned to the TP53 mutant- and wild-type groups, respectively. The survival analyses of both the TP53 mutant- and wild type cohorts yielded similar results. Higher risk scores were associated with worse prognostic outcome (p < 0.05).

### Somatic Mutation Results

Figure 6 illustrates the somatic landscape of the three risk subgroups. Information on the mutation status of each gene in each sample was shown in the waterfall plot, where different colors with specific annotations at the bottom indicated the various types of mutation. The barplot above the legend exhibited the number of mutations. The results showed that CTNNB1 was the most commonly mutated gene in the low-risk group. Tumor protein p53 (TP53) was the most frequently mutated gene in both the moderate- and high-risk groups. Hence, we further grouped the HCC samples into two groups based on the TP53 mutation status. A total of 101 and 241 HCC samples were assigned to the TP53 mutant- and wild-type groups, respectively. The survival analyses of both the TP53 mutant- and wild type cohorts yielded similar results. Higher risk scores were associated with worse prognostic outcome (p < 0.05).

### Association Between Risk Score and Immune Infiltration

We further investigated differences in the degree of immune infiltration in various risk groups. Using the TIMER database, we evaluated the immune cell infiltration in samples from the three aforementioned groups. As shown in Figure 7, we found that the level of immune cell infiltration in macrophages, myeloid dendritic cells, neutrophils, and CD4T cells) was significantly different between the three groups. To further investigate the levels of immune cell infiltration on the gene level, the following eight immune-related genes were selected: sialic acid binding Ig like lectin 15 (SIGLEC15), T cell immunoreceptor with Ig and ITIM domains (TIGIT), CD274, hepatitis A virus cellular receptor 2 (HAVCR2), programmed cell death 1 (PDCD1), cytotoxic T-lymphocyte associated protein 4 (CTLA4), lymphocyte activating 3 (LAG3), and programmed cell death 1 ligand 2 (PDCD1LG2). We compared the expression levels of immune-related genes among the three risk groups. The results are shown in Figure 8. Except for CTLA4, LAG3, and PDCD1, the expression levels of the other five immune-related genes differed significantly between the three groups. Overall, the high-risk group showed significantly higher levels of immune gene expression and immune cell infiltration compared with the other groups.

### Construction and Evaluation of the Nomogram Model

As shown in Table 3, we performed the univariate and multivariate analyses using the SPSS software (version 23.0; IBM Corporation, Armonk, NY, United States) to identify independent prognostic factors predicting OS in patients with HCC. In the training data set, the risk score, T stage, M stage, pTNM stage, and recurrence were identified as independent prognostic factors. The results of the testing and training data sets were similar. Overall, the risk score, T stage, and pTNM stage were identified as independent prognostic factors in both data sets.

Hence, we used age, T stage, N stage, M stage, pTNM stage, recurrence, and risk group as estimated factors in the construction of our model. A nomogram was constructed to estimate the probabilities for three- and 5-year survival. Calibration curves analysis showed that the new nomogram model had good predictive accuracy (Figure 9B). Furthermore, the new nomogram model achieved an area under the curve (AUC) of 0.78 at 5 years, which was better than that of a model without the gene signature (AUC = 0.73) or pTNM stage (AUC = 0.70) (Figure 9C).

### DISCUSSION

The prognosis of patients with HCC varies greatly. The 5-year survival rate after resection of early HCC can be as high as 70%, whereas that of patients with vascular invasion or advanced HCC is markedly lower (Mazzaferro et al., 2011; Grandhi et al., 2016). Therefore, the use of effective prognostic indicators can promptly identify high-risk patients and assist in implementing individualized treatment to improve the prognosis.

In the present study, we found that the AURKA gene was highly expressed in HCC and an independent prognostic risk factor. The GO analysis of AURKA and related genes indicated that these genes are involved in a variety of biological processes, including nuclear chromosome condensation, centromeric regions, mitotic spindle and spindle poles, mitotic nuclear envelope disassembly, and ubiquitin protein ligase activity. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis showed that AURKA and related genes regulate the progression of HCC through multiple pathways, such as the TP53 signaling pathway and FOXO signaling pathway. Numerous studies have demonstrated that the TP53 signaling pathway is involved in the development of a variety of tumors and plays a regulatory role in tumor immunity (Stegh, 2012; Blagih et al., 2020; Muñoz-Fontela et al., 2016). The role of the FOXO

### TABLE 2 | Correlation among risk groups with clinical features in training set.

| Clinical features | Risk group | p value |
|------------------|------------|---------|
|                  | Low Mid High |         |
| Age <50          | 23 12 10    | 0.0339  |
| ≥50              | 112 66 17   |         |
| Gender Female    | 42 29 9     | 0.6639  |
| Male             | 93 49 18    |         |
| T Early (T1+T2)  | 113 52 11   | <0.001  |
| Late (T3+T4)     | 21 26 16    |         |
| N N0             | 90 58 23    | 1       |
| N1+N2            | 1 1 0       |         |
| M M0             | 90 58 23    | 1       |
| M1               | 1 0 0       |         |
| Stage Stage I/II | 106 49 11   | <0.001  |
| Stage III/IV     | 21 24 15    |         |
| Grade G1+G2      | 98 45 9     | 0.0002  |
| G3+G4            | 36 31 18    |         |
| Recurrence Yes   | 72 23 11    | 0.0031  |
| No               | 63 55 16    |         |
| Status Alive     | 96 42 10    | 0.0009  |
| Dead             | 39 38 17    |         |

Frontiers in Genetics | www.frontiersin.org 8 November 2021 | Volume 12 | Article 730732
The signaling pathway in tumors has also received extensive attention (Farhan et al., 2017; Coomans de Brachène and Demoulin, 2016). In general, the enrichment analysis revealed some potential mechanisms and possible pathways of AURKA and its related genes in HCC. In addition, it provided some new ideas for the treatment of patients with HCC.

Among the nine genes associated with AURKA, eight genes (BIRC5, CDC20, PLK1, TPX2, CDK1, CENPA, DLGAP5, and TACC3) were identified as prognostic genes in HCC. Therefore, we used these nine genes to construct the AURKA-related gene signature for the prediction of prognosis of patients with liver cancer. In gastric cancer, high expression of BIRC5 promotes gastric cancer metastasis and is associated with poor prognosis (Zou et al., 2019). However, in lung cancer, high expression of BIRC5 may prolong OS and DFS (Vischioni et al., 2004). In pancreatic cancer, CDC20 can promote tumor cell proliferation and affect the progression of pancreatic cancer (Chang et al., 2012). CDK1 is considered a synthetic target for KRAS-mutated

**FIGURE 6** | Association between somatic mutation and risk score. (A) Overview of somatic mutations in different risk groups. (B) Comparison of AURKA expression between TP53 mutation group and Tp53 wild group. (C) Kaplan-Meier survival analysis of the different risk groups of TP53 mutation. (D) Kaplan-Meier survival analysis of the different risk groups of Tp53 wild.

**FIGURE 7** | Association between risk score and immune infiltration. Macrophage, Myeloid dendritic cell, Neutrophil and CD4+ T cell are significantly high in high risk score groups. *p < 0.05; **p < 0.01; ***p < 0.001.
tumors and has been identified as a prognostic marker in numerous types of cancer (Sung et al., 2014; Costa-Cabral et al., 2016). CENPA plays a key role in cell mitosis. Of note, it is abnormally expressed in tumors and regulates tumor cell activity (Valdivia et al., 2009). DLGAP5 expression is regulated by the ubiquitin-proteasome pathway and participates in tumor cell migration and invasion (Hsu et al., 2004). PLK1 is a key regulator of mitosis and is involved in multiple stages of mitosis (De et al., 2018). Downregulation of PLK1 can inhibit the invasion and metastasis of esophageal cancer cells (Li et al., 2014). In colorectal cancer, PLK1 may promote the growth, invasion, and metastasis of colorectal cancer cells through the PDKI-PLK1-MYC signaling pathway (Tan et al., 2013). TACC3 can activate the Akt/RAS/mitogen-activated protein kinase kinase/extracellular signal–regulated kinase (Akt/RAS/MEK/ERK) signaling pathway to promote the malignant transformation of cells (Burgess et al., 2018). High TACC3 expression has been found in a variety of tumors and is closely related to poor prognosis (Wang et al., 2017). TPX2 is mainly involved in centrosomal maturation and spindle formation (Gruss and Vernos, 2004). In gliomas, MiR-1294 can target TPX2 to inhibit tumor cell proliferation and enhance sensitivity to chemotherapy (Chen et al., 2018)

Several previous studies have successfully constructed multi-gene signatures for risk stratification and prognosis prediction in HCC (Zhou et al., 2020; Ouyang et al., 2020). In this study, we constructed a new nine-gene signature to predict the prognosis of patients with liver cancer. This gene signature was verified using an internal verification data set. Gene signatures can effectively classify patients into high-, moderate-, and low-risk groups. Higher risk scores in the training and validation data sets indicated a poor prognosis in HCC. Finally, we constructed a personalized nomogram based on the risk scores, with a concordance index of 0.78. As an important tumor suppressor gene, TP53 plays a vital role in cell cycle regulation. TP53 mutation is a common mutation in tumors and the most important mutation in liver cancer. This mutation can promote the proliferation, migration, and invasion of tumor cells and increase resistance to drugs (Warren et al., 2013). We divided the HCC cohort of TCGA data set into two groups (TP53 mutation and wild type) and investigated the relationship between the gene signature and these two groups. We found that the gene signature could effectively predict the risk of patients in the TP53 mutation

TABLE 3 | Univariate and multivariate Cox regression analyses of risk factors associated with overall survival.

| HCC cohorts       | Univariable analyses | Multivariable analyses |
|-------------------|----------------------|------------------------|
|                   | p   | HR     | 95.0% CI          | p   | HR     | 95.0% CI          |
| Validation set    | Age | 0.418  | 1.012   | 0.984–1.041       | 0   | 8.257  | 2.168–31.451      |
|                   | Gender | 0.149  | 0.592   | 0.29–1.027        | 0.908  | 1.143  | 0.118–11.109      |
|                   | T stage | 0.04   | 4.017   | 1.925–8.38        | 0.088  | 3.602   | 0.825–15.724      |
|                   | N stage | 0.088  | 3.602   | 2.114–10.017      | 0.003  | 5.318   | 1.778–15.907      |
|                   | M stage | 0.088  | 3.602   | 2.114–10.017      | 0.003  | 5.318   | 1.778–15.907      |
|                   | pTNM, stage | 0.088  | 3.602   | 2.114–10.017      | 0.003  | 5.318   | 1.778–15.907      |
|                   | grade | 0.025  | 3.602   | 2.114–10.017      | 0.003  | 5.318   | 1.778–15.907      |
|                   | recurrence | 0.088  | 3.602   | 2.114–10.017      | 0.003  | 5.318   | 1.778–15.907      |
|                   | Low risk group | 0.088  | 3.602   | 2.114–10.017      | 0.003  | 5.318   | 1.778–15.907      |
|                   | Mid risk group | 0.088  | 3.602   | 2.114–10.017      | 0.003  | 5.318   | 1.778–15.907      |
|                   | High risk group | 0.088  | 3.602   | 2.114–10.017      | 0.003  | 5.318   | 1.778–15.907      |
| Training set      | Age | 0.399  | 1.007   | 0.991–1.023       | 0   | 0.019  | 1.007–3.594       |
|                   | Gender | 0.65   | 0.906   | 0.592–1.387       | 0.028  | 1.968  | 1.078–3.594       |
|                   | T stage | 0.824  | 1.253   | 0.173–9.092       | 0.517  | 1.95   | 0.258–14.738      |
|                   | N stage | 0.012  | 13.563  | 1.763–104.315     | 0.003  | 26.515  | 2.979–236.021     |
|                   | pTNM, stage | 0.17   | 1.349   | 0.88–2.068        | 0   | 0.048  | 1.967–3.851       |
|                   | grade | 0.016  | 1.754   | 1.111–2.769       | 0   | 0.048  | 1.967–3.851       |
|                   | recurrence | 0.207  | 1.193   | 0.575–2.476       | 0   | 0.048  | 1.967–3.851       |
|                   | Low risk group | 0.088  | 3.602   | 2.114–10.017      | 0.003  | 5.318   | 1.778–15.907      |
|                   | Mid risk group | 0.088  | 3.602   | 2.114–10.017      | 0.003  | 5.318   | 1.778–15.907      |
|                   | High risk group | 0.088  | 3.602   | 2.114–10.017      | 0.003  | 5.318   | 1.778–15.907      |
group, thereby effectively classifying these patients into low-, moderate-, and high-risk groups. Numerous recent studies have also confirmed the close relationship between TP53 mutation and tumor immunity (Long et al., 2019; Wu et al., 2020; Sun et al., 2020). Based on the immunoprognostic model established by TP53 mutation, Long et al. found that the levels of T cell follicular helper proteins, T cell regulatory proteins, and macrophages M0 were higher in the high-risk HCC group versus the other groups (Long et al., 2019). Notably, the expression of immune checkpoints CTLA4, PDCD1, and T-cell immunoglobulin mucin family member 3 (TIM3) were also higher in the high-risk group. The investigators suggested that TP53 mutations significantly reduced the immune response in liver cancer.

Immune cell infiltration affects tumor progression. Numerous immunotherapies have been used to regulate immune cells in tumors. Therefore, we investigated the immune cell infiltration in different risk groups. We found that the numbers of macrophages, myeloid dendritic cells, neutrophils, and CD4+ T cells differed significantly in different risk score groups. Higher risk scores were linked to higher numbers of these four types of immune cells. Macrophages play a dual role in the tumor microenvironment, promoting tumor formation and development as well as inhibiting tumor growth (Kim and Bae, 2016). It has been confirmed that the degree of macrophage infiltration in the tumor microenvironment is related to prognosis (Conway et al., 2016). It is currently thought that neutrophils in the tumor microenvironment directly kill or

---

**FIGURE 9** The prognostic nomogram with the risk score in HCC. (A) A nomogram for predicting 3- and 5-year survival possibilities of HCC. (B) The calibration curve of 3-year and 5-year survival. (C) Time-dependent receiver operating characteristic (ROC) curves for gene signature and TNM stage.
stimulate the immune system to inhibit tumor cells and can also promote immune escape of tumor cells (Kim and Bae, 2016). Myeloid dendritic cells mainly play an antigen-presenting role to activate T cells and induce immune responses (Garris and Luke, 2020). CD4+ T cells mainly support CD8+ T toxic lymphocytes and enhance their anti-tumor immune effect (Ghiringhelli et al., 2006). Our results showed differences in the distribution of immune cells in the tumor microenvironment among the different risk groups and revealed the underlying reason for the poor prognosis observed in the high-risk group. To the best of our knowledge, this is the first study to investigate the relationship between AURKA and related genes and tumor immunity.

Immune checkpoint inhibitors are a new approach to the treatment of tumors. Several immune checkpoint inhibitors have been used effectively in the treatment of liver cancer (Liu and Qin, 2019). In this study, we also assessed the relationship between the risk score and immune checkpoints. The results found that the expression of the five immune checkpoints (SIGLEC15, TIGIT, CD274, HAVCR2, and PDCD1LG2) varied in different risk groups and revealed the underlying reason for the poor prognosis observed in the high-risk group. The gene signature based on AURKA successfully classified patients with liver cancer into high-, moderate- and low-risk groups. Hence, the gene signature can be an effective marker for the prognosis of HCC. In addition, the risk score was related to immune cell infiltration and immune checkpoint expression in HCC.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

GW and TD designed the study; RL, ZJ, and WK equally contributed to the literature search, analysis, and writing of the manuscript. SZ supervised this work; All authors read and approved the final manuscript.

REFERENCES

Ai, L., Xu, A., and Xu, J. (2020). Roles of PD-1/pd-L1 Pathway: Signaling, Cancer, and Beyond. Adv. Exp. Med. Biol. 1248, 33–59. doi:10.1007/978-981-15-3266-5_3
Barclay, J., Creswell, J., and León, J. (2018). Cancer Immunotherapy and the PD-1/pd-L1 Checkpoint Pathway. Arch. Esp. Urol. 71, 399–399.
Blaghi, J., Buck, M. D., and Voussen, K. H. (2020). p53, Cancer and the Immune Response. J. Cel. Sci. 133, jcs237453. doi:10.1242/jcs.237453
Briassoulis, P., Chan, F., Savage, K., Reis-Filho, J. S., and Linardopoulos, S. (2007). Aurora-A Regulation of Nuclear Factor-Kb Signaling by Phosphorylation of IκBα. Cancer Res. 67, 1689–1695. doi:10.1158/0008-5472.CAN-06-2272
Burgess, S. G., Mukherjee, M., Sabir, S., Joseph, N., Gutierrez-Caballero, C., Richards, M. W., et al. (2018). Mitotic Spindle Association of TACC3 Requires Aurora-A-dependent Stabilization of a Cryptic a-helm. EMBO. J. 37, e97902. doi:10.15252/emboj.201797902
Camp, R. L., Dollé-Filhart, M., and Kimm, D. L. (2004). X-tile. Clin. Cancer Res. 10, 7252–7259. doi:10.1158/1078-0432.CCR-04-0713
Carmena, M., and Earnshaw, W. C. (2003). The Cellular Geography of Aurora Kinases. Nat. Rev. Mol. Cell. Biol. 4, 842–854. doi:10.1038/nrm1245
Chang, D. Z., Ma, Y., Ji, B., Liu, Y., Hwu, P., Abbruzzese, J. L., et al. (2012). Increased CDC20 Expression Is Associated with Pancreatic Ductal Adenocarcinoma Differentiation and Progression. J. Hematol. Oncol. 5, 15. doi:10.1186/1756-8722-5-15
Chen, C., Song, G., Xiang, J., Zhang, H., Zhao, S., and Zhan, Y. (2017). AURKA and Related Genes Show Better Prognosis and Could Be Potential Therapeutic Targets for HCC. Cancer Lett. 384, 72–80. doi:10.1016/j.canlet.2016.11.008
Chen, H., Liu, L., Li, X., Shi, Y., and Liu, N. (2018). MicroRNA-1294 Inhibits the Proliferation and Enhances the Chemosensitivity of Glioma to Temozolomide via the Direct Targeting of TPX2. Am. J. Cancer Res. 8, 291–301.
Chen, W., Zheng, R., Baade, P. D., Zhang, S., Zeng, H., Bray, F., et al. (2016). Cancer Statistics in China, 2015: CA: A Cancer J. Clinicians 66, 115–132. doi:10.3322/caac.21338
Conway, E. M., Pikor, L. A., Kung, S. H. Y., Hamilton, M. J., Lam, S., Lam, W. L., et al. (2016). Macrophages, Inflammation, and Lung Cancer. Am. J. Respir. Crit. Care Med. 193, 116–130. doi:10.1164/rccm.201508-1545CI
Darvin, C., Joller, T., Rampino, M., Dangel, S., Scheid, F., Taillat, D., et al. (2011). The expression of these immune checkpoints significantly affects tumor prognosis. At present, the use of single immune checkpoint blockers or combinations of these agents has shown good efficacy in different tumors (Darvin et al., 2018).

This study is characterized by several limitations. Firstly, all analyses were based on public databases. The specific mechanisms of AURKA and related genes in liver cancer have not been thoroughly investigated. Furthermore, the gene signature was associated with immune infiltration and immune checkpoint expression in HCC and affected the prognosis of patients with this disease. Further studies are needed to examine the value of gene signatures in immune invasion and prognosis in HCC.

In summary, we found that nine AURKA-related genes with prognostic value can be used as diagnostic markers for liver cancer. The gene signature based on AURKA successfully classified patients with liver cancer into high-, moderate- and low-risk groups. Hence, the gene signature can be an effective marker for the prognosis of HCC. In addition, the risk score was related to immune cell infiltration and immune checkpoint expression in HCC.
Coomas de Brachène, A., and Demoulin, J.-B. (2016). FOXO Transcription Factors in Cancer Development and Therapy. Cell. Mol. Life Sci. 73, 1159–1172. doi:10.1007/s00018-015-2112-y
Costa-Cabrál, S., Brough, R., Konde, A., Aarts, M., Campbell, J., Marinari, E., et al. (2016). CDK1 Is a Synthetic Lethal Target for KRAS Mutant Tumours. PloS. One. 11, e0149099. doi:10.1371/journal.pone.0149099
Darvin, P., Toor, S. M., Sasidharan Nair, V., and Ellkord, E. (2018). Immune Checkpoint Inhibitors: Recent Progress and Potential Biomarkers. Exp. Mol. Med. 50, 1–11. doi:10.1038/s41276-018-0191-1
Das, M., Zhu, C., and Kuchroo, V. K. (2017). Tim-3 and its Role in Regulating Autoimmune Diseases. Frontiers in Immunology. 8, 1654. doi:10.3389/fimmu.2017.01654
de C à s, M., Zhu, C., and Kuchroo, V. K. (2017). Tim-3 and its Role in Regulating Autoimmune Diseases. Front. Immunol. 9, 2909. doi:10.3389/fimmu.2018.02999
Long, J., Wang, A., Bai, Y., Lin, J., Yang, X., Wang, D., et al. (2019). Development and Validation of a TP53-Associated Immune Prognostic Model for Hepatocellular Carcinoma. EBioMedicine. 42, 363–374. doi:10.1016/j.ebiom.2019.03.022
Lu, L., Han, H., Tian, Y., Li, W., Zhang, J., Feng, M., et al. (2015). Aurora Kinase A Mediates C-Myc’s Oncogenic Effects in Hepatocellular Carcinoma. Mol. Carcinog. 54, 1467–1479. doi:10.1002/mc.22223
Mazaferro, V., Bhoori, S., Sposito, C., Bongini, M., Langer, M., Miceli, R., et al. (2011). Milan Criteria in Liver Transplantation for Hepatocellular Carcinoma: An Evidence-Based Analysis of 15 Years of Experience. Liver. Transpl. 17, S44–S57. doi:10.1002/lst.22365
Mering, C. V., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. (2003). STRING: A Database of Predicted Functional Associations Between Proteins. Nucleic Acids Res. 31, 258–261. doi:10.1093/nar/gkg034
Muñoz-Fontela, C., Mandinova, A., Aaronson, S. A., and Lee, S. W. (2016). Emerging Roles of P53 and Other Tumour-Suppressor Genes in Immune Regulation. Nat. Rev. Immunol. 16, 741–750. doi:10.1038/nri.2016.99
Ouyang, G., Yi, B., Chan, X., and Chen, X. (2020). A Robust Twelve-Signature Probe Set for Prognosis Prediction of Hepatocellular Carcinoma. Cancer Cell. Int. 20, 207. doi:10.1186/s12933-020-02194-9
Reiter, R., Gaiss, P., Jüttig, U., Steuer-Vogt, M. K., Pickhard, A., Bink, K., et al. (2006). Aurora Kinase A Messenger RNA Overexpression Is Correlated with Tumor Progression and Survival in Head and Neck Squamous Cell Carcinoma. Clin. Cancer Res. 12, 5136–5141. doi:10.1158/1059-044X.CCR-05-1650
Stegh, A. H. (2012). Targeting the P53 Signaling Pathway in Cancer Therapy - The Promises, Challenges and Perils. Expert Opin. Ther. Targets 16, 67–83. doi:10.1517/17512433.2012.1171
Sun, H., Liu, S.-Y., Zhou, J.-Y., Xu, J.-T., Zhang, H.-K., Yan, H.-H., et al. (2020). Specific TP53 Subtype as Biomarker for Immune Checkpoint Inhibitors in Lung Adenocarcinoma. EBioMedicine. 60, 102990. doi:10.1016/j.ebiom.2020.102990
Sung, H., Lee, P.-R., Wu, P.-R., Yu, C.-T. R., and Huang, C.-Y. F. (2004). Fbx7 Functions as a Regulated Protein (HURP) Proteolysis by a Proline-Rich Region. J. Cell. Biol. 166, 949–955. doi:10.1083/jcb.200312112
Hartke, J., Johnson, M., and Ghabril, M. (2017). The Diagnosis and Treatment of Hepatocellular Carcinoma. Semin. Diagn. Pathol. 34, 153–159. doi:10.1053/j.sdpd.2016.12.011
Hsu, J.-M., Lee, Y.-C. G., Yu, C.-T. R., and Huang, C.-Y. F. (2004). Cutting Edge: TIGIT Has T Cell-Intrinsic Inhibitory Functions. J. Immunol. 172, 5360–5362. doi:10.4049/jimmunol.172.10.5360
Jeng, Y.-M., Peng, S.-Y., Lin, C.-Y., and Hsu, H.-C. (2004). Overexpression and Amplification of Aurora-A in Hepatocellular Carcinoma. Clin. Cancer Res. 10, 2065–2071. doi:10.1158/1078-0432.CCR-03-1911
Joller, N., Hafer, J. P., Reynedal, B., Kassam, N., Spoerl, S., Levin, S. D., et al. (2011). Cutting Edge: TIGIT Has T Cell-Intrinsic Inhibitory Functions. J. Immunol. 186, 1338–1342. doi:10.4049/jimmunol.1003081
Katayama, H., Sasaki, K., Kawai, H., Yuan, Z.-M., Bondurak, J., Suzuki, F., et al. (2004). Phosphorylation by aurora Kinase A Induces Mdm2-Mediated Distabilization and Inhibition of P53. Nat. Genet. 36, 55–62. doi:10.1038/ng1279
Kim, J., and Bae, J.-S. (2016). Tumor-Associated Macrophages and Neutrophils in Tumor Microenvironment. Mediators Inflamm. 2016, 1–11. doi:10.1155/2016/6058147
Li, C., Chen, J., Zhang, K., Feng, R., Wang, R., and Chen, L. (2015). Progress and Prospects of Long Non-Coding RNAs (IncRNAs) in Hepatocellular Carcinoma. Cell. Physiol. Biochem. 36, 423–434. doi:10.1159/0004030109
Li, C., Zhou, X., Wang, Y., Jing, S., Yang, C., Sun, G., et al. (2014). miR-210 Regulates Esophageal Cancer Cell Proliferation by Inducing G2/M Phase Cell Cycle Arrest Through Targeting PLK1. Mol. Med. Rep. 10, 2099–2104. doi:10.3892/mmr.2014.2416
Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., et al. (2017). TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer Res. 77, e108–e110. doi:10.1158/0008-5472.CAN-17-0307
Lindon, C., Grant, R., and Min, M. (2016). Ubiquitin-Mediated Degradation of Aurora Kinases. Front. Oncol. 5, 307. doi:10.3389/fonc.2015.00307
Liu, X., and Qin, S. (2019). Immune Checkpoint Inhibitors in Hepatocellular Carcinoma: Opportunities and Challenges. Oncol. 24, S3–S10. doi:10.1634/theoncologist.2019-01-S1-001
Locy, H., de Mey, S., de Mey, W., De Ridder, M., Thielemans, K., and Maenhout, S. K. (2018). Immunomodulation of the Tumor Microenvironment: Turn Foe into Friend. Front. Immunol. 9, 2909. doi:10.3389/fimmu.2018.02999
Matoz-Matas, R., Siddle, C., and Llanos, H. (2016). Aurora Kinase A Mediates C-Myc’s Oncogenic Effects in Hepatocellular Carcinoma. Mol. Carcinog. 54, 1467–1479. doi:10.1002/mc.22223
Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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

Copyright © 2021 Liu, Jiang, Kong, Zheng, Dai and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.