High Expression of SRD5A3 in Human Hepatocellular Carcinoma (HCC) Predicts Poor Prognosis: A Study Based on TCGA Data

Jing Xue  
First Clinical Medical College Affiliated to Beijing University of Chinese Medicine

Xianzhao Yang  
Dongzhimen Hospital Affiliated to Beijing University of Chinese Medicine

Feng Jiang (jiangfeng@bucm.edu.cn)  
Dongzhimen Hospital Affiliated to Beijing University of Chinese Medicine

Research Article

Keywords: SRD5A3, human hepatocellular carcinoma, prognosis, bioinformatic analysis

Posted Date: December 2nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1118583/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background:** Hepatocellular carcinoma (HCC) is the fourth leading cause of cancer-related death worldwide. Steroid 5 alpha-reductase 3 (SRD5A3) was reported to be up-regulated in many types of cancer. However, its expression and role in HCC remains to be elucidated. We aim to evaluate the significance of SRD5A3 expression in HCC by using analysis of a public dataset from The Cancer Genome Atlas (TCGA).

**Methods:** The relationship between clinical pathologic features and SRD5A3 were analyzed with the Kolmogorov-Smirnov test and the logistic regression. Cox regression and the Kaplan-Meier method were used to assess the clinicopathologic characteristics associated with overall survival (OS) in TCGA patients. In addition, GSEA was used to predict potential hallmarks associated with different expression of SRD5A3 on transcriptional sequences from TCGA database.

**Results:** SRD5A3 was highly expressed in HCC tumor tissue compared to normal tissue. A total of 184 upregulated DEGs (differentially expressed genes) and 58 downregulated DEGs were identified between high expression and low expression of SRD5A3. Among them, 22 hub genes mainly belonging to the keratin and MUC family demonstrated by connectivity degree in the PPI network were screened out. Kaplan-Meier method showed that HCC patients in the high SRD5A3 expression group had poorer overall survival (OS, HR=2.26(1.58-3.24), p<0.001). In addition, cell cycle mitotic, cell cycle checkpoints, mitotic nuclear division, Q-glycan processing, protein O-linked glycosylation were differentially enriched in the high SRD5A3 expression phenotype pathway. In addition, SRD5A3 expression level has significant correlations with infiltrating levels of Th17 (R = -0.238, p < 0.001), Cytotoxic cells (R = -0.234, p < 0.001) and Th2 cells (R = 0.258, p < 0.001) in HCC.

**Conclusions:** High expression of SRD5A3 was significantly correlated with poor prognosis in HCC patients. It may be a potential biomarker in HCC.

**Background**

According to 2018 statistics from the World Health Organization, primary liver cancer is the sixth most common cancer and the fourth in mortality worldwide, causing about 840,000 new cases and more than 780,000 deaths annually [1]. Of all liver cancer cases, >90% are hepatocellular carcinomas (HCCs) [2]. Patients are often diagnosed with liver cancer in advanced stages, contributing to its poor prognosis [3]. In the past decade, there have been improvements in non-drug therapies and drug therapies for HCC treatment. Surgical resection is the primary treatment for primary hepatocellular carcinoma. However, since hepatocellular carcinoma usually occurs in the context of cirrhosis, residual liver resection still presents a specific tumor microenvironment, and recurrence becomes the most serious problem. In recent years, cancer immunotherapy has become a research hotspot by arousing the immune system to respond to developing cancer, but its clinical application is still limited, and clinical trials for molecular therapies targeting different aspects of the hepatocellular carcinogenesis cascade are under way. To provide a
highly accurate prediction of a patient's survival and/or response to individualized treatment therapy, new biomarkers has been conducted among patients.

Studies have confirmed that steroid 5α reductase (SRD5A3), as an important component of steroid metabolic cascade, irreversibly converts testosterone (T) to dihydrotestosterone (DHT), and binds to androgen receptor (AR) to form DHT-AR complex that specifically acts on AR response elements and promotes the transcription of downstream target genes. Such as the Wnt/β-catenin signaling pathway, CD24, and vascular endothelial growth factor (VEGF), promote tumor proliferation, metastasis and angiogenesis [4]. To date, three types of human 5α-steroid reductase enzymes have been reported (types 1, 2 and 3). Both SRD5A1 and SRD5A2 are expressed in human hepatocytes, and SRD5A1 alone is expressed in adipocytes [5]. Steroid 5 alpha-reductase 3 (SRD5A3) mainly distributed in the brain, retina, heart, liver and other tissues, a protein-coding gene that irreversibly reduces steroid hormones and eventually becomes the corresponding metabolite with biological activity [6]. Recent studies have found that it is overexpressed in a variety of cancers. SRD5A3 overexpressed specifically in HRPC cells and played important roles in HRPC growth and progression. is a promising molecular target for prostate cancer therapy [7,8]. In addition, SRD5A3 is validated as independent prognostic markers of EC and Triple-Negative Breast Cancer [9,10]. However, previous studies have focused on particular aspects of SRD5A3 in other diseases and thus lack a comprehensive exploration and description.

Thus, the objective of this study is to evaluate the prognostic value of SDR5A3 expression in HCC based on data obtained from TCGA. We investigated the expression levels of SRD5A3 and correlated it with survival in HCC patients, in the hope that our findings could be helpful for understanding of pathological process and aggressiveness in HCC. Through immune infiltration analysis, gene enrichment analysis and protein-protein interaction (PPI) network, we further investigated the important role of SRD5A3 in the immune microenvironment and cell cycle, and further explored the potential value of SRD5A3 in immunotherapy and chemotherapeutic resistance therapy related to cell cycle mechanisms. It is expected to provide some guidance for the determination of new targets for precise treatment of HCC.

Materials And Methods

Data acquisition

SRD5A3 expression and clinical data of TCGA pan-cancer data and GTEx were obtained from the UCSC Xena database [22] (https://xenabrowser.net/datapages/). The clinical data for the HCC cohort were obtained from the data from TCGA (https://cancergenome.nih.gov/). Patients who have insufficient clinicopathological data or "0" gene expression values were not included. Since the data come from these online databases, it is not necessary to get the study approval by the ethics committee. The RNA-Seq gene expression FPKM (Fragments Per Kilobase per Million) of 371 cases with HCC and clinical data were retained and further analyzed (Table 1). The HTSeq-FPKM data were transformed to TPM (transcription per million reads) for the following analyses.
Analysis of differentially expressed genes (DEGs) for HCC between high and low SRD5A3 expression groups.

According to the median expression value of AAA in TCGA XXX samples, patients were classified into low- and high-expression groups. Expression profiles (HTSeq-TPM) were compared between high and low SRD5A3 expression groups to identify differentially expressed genes (DEGs) within the limma Package software by using R [23]. Genes with |logFC|>1.5 and FDR<0.05 were considered as DEGs.

Functional Enrichment analysis

The Metascape database (http://metascape.org/) is a free and credible gene-list analysis online tool for high-throughput functional analysis of genes. It can realize habitual and understand common and unique pathways within a group of orthogonal target-discovery studies. Metascape was applied to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. GO analysis is a common useful method for annotating genes and gene products and for identifying molecular function (MF), biological process (BP) and cellular component (CC) attributes for high-throughout genome or transcriptome data. KEGG is a collection of databases used for systematic analysis of gene functions and associating related gene sets with their pathways. In this study, we assessed the functions of SRD5A3 and their co-expression genes. The P value was set as 0.01, and the enrichment factor of >1.5 and minimum count of 3 were considered significant. We then selected a subset of representative terms from this cluster and converted them into a network layout. Molecular Complex Detection (MCODE) was supposed to recognize closely related protein components.

Gene Set Enrichment Analysis

GSEA was used to predict potential hallmarks associated with different expression of SRD5A3 on transcriptional sequences from TCGA database. A permutation test with 10000 times was used to identity the significantly changed pathways. Adj.P less than 0.05 and FDR less than 0.25 were identified as significantly related genes. Statistical analysis and graphical plotting were conducted using R package clusterProfiler (3.8.0) [24].

Immune infiltration analysis by ssGSEA

The immune infiltration analysis of HCC was done by ssGSEA (single-sample gene Set Enrichment Analysis) method using GSVA package [25] in R (3.8.0) for 24 types of immune cells in tumor samples. Based on the signature genes of the 24 types immunocyte in the literature [26], the relative enrichment score of every immunocyte was quantified from gene expression profile for each tumor sample. The correlation between SRD5A3 and these immune cells was analyzed by Spearman correlation, and the
infiltration of immune cells between the high and low-expression groups of SRD5A3 was analyzed by the Wilcoxon rank sum test.

**Statistical analysis**

R (v.3.8.0) software was used to perform the statistical analysis. Because of normal distribution, the differences in the SRD5A3 expression levels between normal tissues and tumors in HCC cohorts were performed through the Wilcoxon signed-rank test. The Kolmogorov-Smirnov test and logistic regression were used to analyze the relationship between SRD5A3 expression and clinical pathologic characters. In this study, we compared SRD5A3 expression in patients with HCC tissues and adjacent normal tissues, and evaluated the discrimination ability of SRD5A3 in HCC by ROC analysis using pROC package. P-value < 0.05 was set up as the cut-off criterion.

**Results**

**Pan-cancer and HCC SRD5A3 expression analysis**

We first assessed SRD5A3 expression in pan-cancer data from TCGA and GTEx. The data revealed that SRD5A3 expression was significantly higher in 29 types of tumors than those corresponding normal tissues, including ACC, BLCA, BRCA, CESC, CHOL, COAD, DLBC, ESCA, GBM, KICH, KIRC, KIPPA, LIHC, LGG, LUAD, LUSC, OV, PAAD, PRAD, READ, SARC, SKCM, STAD, TGCT, THCA, THYM, UCEC, UCS (all P<0.05). (Fig.1a). We further compared the expression of SRD5A3 in GTEX combined with TCGA normal samples and TCGA hepatocellular carcinoma samples, 50 paracancerous and 371 HCC samples from TCGA, 50 HCC samples and their corresponding paired paracancerous samples from TCGA. Finally, it was found that SRD5A3 was significantly overexpressed in HCC samples, and the results were statistically significant (P<0.001). (Fig.1b-d). In addition, the ROC curve with an area under the receiver operating characteristic curve value of 0.849 (95% CI 0.804 to 0.895, p< 0.001), which showed high diagnostic accuracy of SRD5A3 in HCC, as is shown in (Fig.1e).

**Differentially expressed genes (DEGs) analysis**

We performed a differentially expressed genes (DEGs) analysis using TCGA cohort data, patients with HCC were divided into the high expression group and the low expression group according to the expression level of SRD5A3. A total of 242 differentially expressed genes were found using screening, 184 were highly expressed and 58 were lowly expressed, as is shown in Fig.2a. We constructed a gene expression heatmap to get an overview for the top 10 genes with the greatest differences in expression in HCC (Fig.2c).
Association with SRD5A3 expression and clinicopathologic variables

As shown in Table 1, 371 primary tumors with both clinical and gene expression data were downloaded from TCGA data. There were 250 male and 121 female with median age 61 years in the cohort, of the 371 specimens. Correlation analysis showed that SRD5A3 protein expression was significantly associated with race (p = 0.001), T stage (p < 0.001), pathologic stage (p < 0.001), tumor status (p = 0.008), and residual tumor (p = 0.017). No correlation was found between SRD5A3 expression and other clinicopathologic features, detail information was showed in table 1. Univariate analyses using logistic regression revealed that up-regulation of SRD5A3 in HCC is significantly associated with race (p < 0.001), BMI (p = 0.049), pathologic stage (p < 0.001), tumor status (p = 0.006) and residual tumor (p = 0.016). (Table 2).

As shown in Fig 3, the higher expression of SRD5A3 correlated significantly with gender (p<0.01), tumor status (p<0.01), residual tumor (p<0.01), race(Asian vs. White, p<0.01), T stage (T1 vs. T2, p<0.001;T1 vs. T3, p<0.05), and pathologic tumor (stage T vs. stage T, p<0.01; stage T vs. stage T, p<0.01).

Table 1 Relationship between SRD5A3 expression and clinicopathological characteristics of patients [n (%)]
| Characters             | Low expression of SRD5A3 | High expression of SRD5A3 | P value |
|------------------------|--------------------------|---------------------------|---------|
| Gender (%)             |                          |                           |         |
| Female                 | 52(28.0%)                | 69(37.3%)                 | 0.071‡  |
| Male                   | 134(72.0%)               | 116(62.7%)                |         |
| Race (%)               |                          |                           |         |
| Asian                  | 96(53.6%)                | 62(34.4%)                 | 0.001*  |
| Black or African American | 7(3.9%)              | 10(5.6%)                   |         |
| White                  | 76(42.5%)                | 108(60.0%)                |         |
| Age (median [IQR])     | 61.00[51.00,68.00]       | 61.00[52.00,69.00]        | 0.673§  |
| Height (median [IQR])  | 168.00[161.00,174.00]    | 167.00[160.25,174.00]     | 0.610§  |
| Weight (median [IQR])  | 68.00[59.50,79.00]       | 72.00[58.00,85.00]        | 0.344§  |
| BMI (median [IQR])     | 24.01[21.90,27.54]       | 25.28[21.45,29.75]        | 0.296§  |
| T stage (%)            |                          |                           |         |
| T1                     | 112(60.5%)               | 69(37.7%)                 | <0.001* |
| T2                     | 33(17.8%)                | 61(33.3%)                 |         |
| T3                     | 34(18.4%)                | 46(25.1%)                 |         |
| T4                     | 6(3.2%)                  | 7(3.8%)                   |         |
| N stage (%)            |                          |                           |         |
| N0                     | 136(98.6%)               | 116(98.3%)                | 1.000‖  |
| N1                     | 2(1.4%)                  | 2(1.7%)                   |         |
| M stage (%)            |                          |                           |         |
| M0                     | 141(99.3%)               | 125(97.7%)                | 0.348‖  |
| M1                     | 1(0.7%)                  | 3(2.3%)                   |         |
| Pathologic stage (%)   |                          |                           |         |
| Stage I                | 108(60.3%)               | 63(37.5%)                 | <0.001* |
| Stage II               | 33(18.4%)                | 53(31.5%)                 |         |
| Stage III              | 36(20.1%)                | 49(29.2%)                 |         |
| Stage IV               | 2(1.1%)                  | 3(1.8%)                   |         |
| Tumor status (%)       |                          |                           |         |
| Tumor free             | 115(64.2%)               | 86(49.7%)                 | 0.008*  |
|                                | With tumor | 87(50.3%) | 0.017* |
|--------------------------------|------------|-----------|--------|
| Residual tumor (%)             | R0         | 173(97.7%)| 151(91.5%) |
|                                | R1         | 4(2.3%)   | 13(7.9%)  |
|                                | R2         | 0(0.0%)   | 1(0.6%)   |
| Histologic grade (%)           | G1         | 26(14.0%) | 29(16.1%) |
|                                | G2         | 88(47.3%) | 89(49.4%) |
|                                | G3         | 63(33.9%) | 59(32.8%) |
|                                | G4         | 9(4.8%)   | 3(1.7%)   |
| Child-Pugh grade (%)           | A          | 125(93.3%)| 92(87.6%) |
|                                | B          | 8(6.0%)   | 13(12.4%) |
|                                | C          | 1(0.7%)   | 0(0.0%)   |
| Fibrosis ishak score (%)       | 0          | 35(32.1%) | 39(37.9%) |
|                                | 1/2        | 18(16.5%) | 13(12.6%) |
|                                | 3/4        | 16(14.7%) | 12(11.7%) |
|                                | 5/6        | 40(36.7%) | 39(37.9%) |
| Adjacent hepatic tissue inlamination (%) | Mild | 54(42.5%) | 45(42.1%) |
|                                | None       | 62(48.8%) | 55(51.4%) |
|                                | Severe     | 11(8.7%)  | 7(6.5%)   |
| Vascular invasion (%)          | No         | 112(68.3%)| 94(62.3%) |
|                                | Yes        | 52(31.7%) | 57(37.7%) |
| TP53 status (%)                | Mut        | 53(29.1%) | 49(27.8%) |
|                                | WT         | 129(70.9%)| 127(72.2%)|
| Total bilirubin(mg/dl) (median [IQR]) | 0.60[0.50,0.90] | 0.70[0.50,1.02] |
| Albumin(g/dl) (median [IQR])   | 4.00[3.50,4.30] | 4.00[3.30,4.40] |
| AFP (ng/ml)                    | 11.00[4.00,322.25] | 22.50[5.00,231.50] | 0.273$
Prothrombin time (median [IQR])

1.10[1.00,8.70] 1.10[1.00,10.20] 0.457§

*Statistically significant; § Wilcoxon rank sum test; † Fisher exact test; ‡ χ2 test

**Table 2** SRD5A3 expression associated with clinical pathological characteristics (logistic regression)

| Characteristics | Total(N) | Odds Ratio (OR) | P value |
|-----------------|----------|-----------------|---------|
| Gender (Female vs. Male) | 371 | 1.53(0.99-2.38) | 0.056 |
| Race (Asian vs. Black or African American & White) | 359 | 0.45(0.30-0.69) | <0.001 |
| Age (<=60 vs. >60) | 370 | 0.90(0.60-1.35) | 0.603 |
| Height (< 170 vs. >=170) | 339 | 1.05(0.68-1.61) | 0.842 |
| Weight (<=70 vs. >70) | 344 | 0.68(0.44-1.04) | 0.073 |
| BMI (<=25 vs. >25) | 335 | 0.65(0.42-1.00) | 0.049 |
| T stage (T1&T2 vs. T3&T4) | 368 | 0.68(0.42-1.08) | 0.106 |
| N stage (N0 vs. N1) | 256 | 0.85(0.10-7.20) | 0.875 |
| M stage (M0 vs. M1) | 270 | 0.30(0.01-2.34) | 0.294 |
| Pathologic stage (Stage I vs. Stage II &Stage III &Stage IV) | 347 | 0.39(0.25-0.61) | <0.001 |
| Tumor status (Tumor free vs. With tumor) | 352 | 0.55(0.36-0.84) | 0.006 |
| Residual tumor (R0 vs. R1&R2) | 342 | 0.25(0.07-0.71) | 0.016 |
| Histologic grade (G1&G2 vs. G3&G4) | 366 | 1.20(0.79-1.84) | 0.397 |
| Child-Pugh grade (A vs. B&C) | 239 | 0.51(0.20-1.23) | 0.138 |
| Fibrosis ishak score (0 vs. 1/2&3/4&5/6) | 212 | 1.29(0.73-2.28) | 0.380 |
| Adjacent hepatic tissue inflammation (Mild vs. None &Severe) | 234 | 0.98(0.58-1.65) | 0.943 |
| Vascular invasion (No vs. Yes) | 315 | 0.77(0.48-1.22) | 0.261 |
| TP53 status (Mut vs. WT) | 358 | 0.94(0.59-1.49) | 0.789 |
| Total bilirubin(mg/dl) (<2 vs. >=2) | 301 | 0.56(0.20-1.49) | 0.250 |
| Albumin(g/dl) (<3.5 vs. >=3.5) | 297 | 1.51(0.88-2.60) | 0.137 |
| AFP (ng/ml) (<400 vs. >400) | 278 | 1.21(0.70-2.14) | 0.497 |
| Prothrombin time (<=4 vs. >4) | 294 | 0.72(0.44-1.19) | 0.203 |
Clinical value of SRD5A3 in prognosis.

To confirm the correlation between the SRD5A3 expression and the prognosis of HCC, survival rates between the high and low SRD5A3 level groups were compared. The Kaplan–Meier survival analysis found that HCC patients in the high SRD5A3 expression group had poorer overall survival (OS, HR=2.26(1.58-3.24), p<0.001) (Fig. 4). T stage (p < 0.001), M stage (p = 0.018), Pathologic stage (p < 0.001), Tumor status (p < 0.001) and SRD5A3 expression were included in Cox proportional risk regression model for multivariate analyses. Multivariate analyses showed that SRD5A3 remained independently associated with overall survival (HR=2.382(1.447-3.920), p<0.05), along with tumor status, TP53 status, T stage (all p <0.05), as is shown in Table 3.

Based on Cox proportional risk regression model, tumor status, TP53 status, T stage and the expression of SRD5A3 were included in the nomogram (Fig.5a). The C-index of the prognostic model was 0.692(95%CI:0.666-0.719). We constructed calibration plots evaluating the agreement between the predicted and the actual OS for the prognosis model, and the results showed that the predicted results of the nomogram were reliable (Fig. 5b).

Table3 Univariate and multivariate analyses of overall survival in HCC patients from TCGA.
| Characteristics                                      | Univariate analysis HR (95% CI) | P value  | Multivariate analysis HR (95% CI) | P value |
|------------------------------------------------------|--------------------------------|----------|----------------------------------|---------|
| Gender (Female vs. Male)                             | 1.225(0.860-1.746)             | 0.260    |                                  |         |
| Race (Asian vs. Black or African American &White)    | 0.760(0.525-1.101)             | 0.146    |                                  |         |
| Age (<=60 vs. >60)                                   | 0.802(0.565-1.136)             | 0.214    |                                  |         |
| Height (< 170 vs. >=170)                             | 0.828(0.570-1.201)             | 0.319    |                                  |         |
| Weight (<=70 vs. >70)                                | 1.091(0.762-1.562)             | 0.634    |                                  |         |
| BMI (<=25 vs. >25)                                   | 1.232(0.843-1.775)             | 0.289    |                                  |         |
| T stage (T1&T2 vs. T3&T4)                            | 0.394(0.277-0.560)             | <0.001   | 0.499(0.280-0.888)               | 0.018   |
| N stage (N0 vs. N1)                                  | 0.499(0.122-2.037)             | 0.333    |                                  |         |
| M stage (M0 vs. M1)                                  | 0.248(0.078-0.789)             | 0.018    | 0.712(0.168-3.021)               | 0.645   |
| Pathologic stage (Stage I vs. Stage II & Stage III & Stage IV) | 0.482(0.330-0.705)             | <0.001   | 0.763(0.404-1.442)               | 0.405   |
| Tumor status (Tumor free vs. With tumor)             | 0.424(0.291-0.617)             | <0.001   | 0.521(0.310-0.875)               | 0.014   |
| Residual tumor (R0 vs. R1&R2)                        | 0.637(0.322-1.258)             | 0.194    |                                  |         |
| Histologic grade (G1&G2 vs. G3&G4)                   | 0.893(0.623-1.281)             | 0.539    |                                  |         |
| Child-Pugh grade (A vs. B&C)                         | 0.619(0.305-1.254)             | 0.183    |                                  |         |
| Fibrosis ishak score (0&1/2 vs. 3/4&5/6)             | 1.320(0.794-2.197)             | 0.285    |                                  |         |
| Adjacent hepatic tissue inflammation (None vs. Mild &Severe) | 0.815(0.501-1.325)             | 0.409    |                                  |         |
| Vascular invasion (No vs. Yes)                       | 0.742(0.490-1.124)             | 0.159    |                                  |         |
| Characteristics                               | Univariate analysis |                      |                      |                      |                      |
|----------------------------------------------|---------------------|----------------------|----------------------|----------------------|----------------------|
|                                              | HR (95% CI)         | P value              | HR (95% CI)         | P value              |                      |
| TP53 status (Mut vs. WT)                     | 1.434 (0.972-2.115) | 0.069                | 1.818 (1.083-3.050) | 0.024                |                      |
| Total bilirubin (mg/dl) (<2 vs. >=2)         | 0.858 (0.347-2.118) | 0.740                |                      |                      |                      |
| Albumin (g/dl) (<3.5 vs. >=3.5)              | 1.085 (0.665-1.771) | 0.743                |                      |                      |                      |
| AFP (ng/ml) (<=400 vs. >400)                 | 0.947 (0.579-1.548) | 0.827                |                      |                      |                      |
| Prothrombin time (<=4 vs. >4)                | 0.752 (0.496-1.140) | 0.179                |                      |                      |                      |
| SRD5A3 (High vs. Low)                        | 2.261 (1.580-3.236) | <0.001               | 2.382 (1.447-3.920) | <0.001               |                      |

**Gene sets enriched in SRD5A3 expression phenotype and protein-protein interaction**

The functions of SRD5A3 were predicted by analyzing GO and KEGG in Metascape. The top 20 GO enrichment items were classified into three functional groups: biological process group (11 items), cellular component group (3 items), and molecular function group (6 items) (Fig. 6a, b). Regarding to the KEGG pathway analysis as shown in Fig. 6c. Mineral absorption, gastric cancer, Salivary secretion, Protein digestion and absorption, parathyroid hormone synthesis, secretion and action were the significant metabolic pathways.

In addition, to better understand the potential biological function of SRD5A3 in HCC, we performed a Metascape protein–protein interaction enrichment analysis. The protein–protein interaction network and MCODE components identified in the gene lists are shown in Fig. 6d. The four most significant MCODE components were extracted from the protein–protein interaction network. After pathway and process enrichment analysis was independently applied to each MCODE component, the results showed that biological function was mainly related to comification, intermediate filament, keratinization, O-glycan processing, protein O-lined glycosylation, macromolecule glycosylation, G protein-coupled receptor binding, cellular response to lipid, anchored component of membrane. A single gene may be correlated with multiple target genes, which may form functionally interactive modules involved in regulation of SRD5A3 in HCC.
connected network, we obtained 22 hub genes, including KRT19, KRT1, KRT6C, KRT80, KRT4, KRT12, MUC5B, MUC21, MUC5AC, MUC1, CHST4, CEACAM7, PRSS21, YPD68, GP2, SAA1, PENK, SSTR5, CXCL5, FGF23, FDF5, CALML3, as is shown in Fig. 6e. These hub genes mainly belong to the keratin and MUC family.

To identify signaling pathways that are differentially activated in HCC, we further conducted GSEA between low and high SRD5A3 expression data sets. Using MSigDB Collection (c2.cp.v7.0.symbols.gmt and c5.all.v7.0.symbols.gmt), we selected the most significantly enriched signaling pathways based on their normalized enrichment score (NES) (Fig. 7, Table 4). Due to the limited space, only six pathways of high and expression are listed here. GSEA identified cell cycle mitotic (Fig. 7a), cell cycle checkpoints (Fig. 7b), mitotic nuclear division (Fig. 7c), O-glycan processing (Figure 7d), mitotic prometaphase (Fig. 7e), protein O-linked glycosylation (Fig. 7f) were differentially enriched in the high SRD5A3 expression phenotype pathway.

Table 4. GSEA analysis of SRD5A3 related enriched gene set with high expression of SRD5A3

| MsigDB collection | Gene set name                              | NES   | p.adjust | FDR  |
|-------------------|--------------------------------------------|-------|----------|------|
| c2.cp.v7.0.symbols.gmt | REACTOME_CELL_CYCLE_MITOTIC                   | 3.02  | 0.016    | 0.01 |
| c5.all.v7.0.symbols.gmt | REACTOME_CELL_CYCLE_CHECKPOINTS        | 2.998 | 0.016    | 0.01 |
|                   | GO/MITOTIC_NUCLEAR_DIVISION               | 2.983 | 0.013    | 0.008|
|                   | GO_O_GLYCAN_PROCESSING                    | 2.979 | 0.013    | 0.008|
|                   | REACTOME/MITOTIC_PROMETAPHASE             | 2.918 | 0.016    | 0.01 |
|                   | GO_PROTEIN_O_LINKED_GLYCOSYLATION          | 2.897 | 0.013    | 0.008|

NES: normalized enrichment score; FDR: false discovery rate. Gene sets with FDR<0.25 and p.adjust<0.05 are considered as significant.

**Relationship between SRD5A3 expression and tumor-infiltrating immune cells**

Analysis and characterisation of proteins and genes involved in cancer development at the molecular level, could add to our knowledge of potential prognostic factors [11]. Numerous studies had now documented a link between the immune infiltrate in several human carcinoma types and prognosis and response to therapy [12]. Therefore, we tried to find whether SRD5A3 expression was associated with immune infiltration in HCC. Spearman correlation was employed to show the association between the expression level (TPM) of SRD5A3 and immune cell infiltration level quantified by GSEA in the HCC tumor microenvironment. Additionally, our findings strongly underline the significant role of SRD5A3 in immune infiltration. SRD5A3 was negatively correlated Th17(R = -0.238, p < 0.001) (Fig. 8a), Cytotoxic cells (R =
-0.234, p < 0.001) (Fig. 8b), whereas it was positively correlated with Th2 cells (R = 0.258, p < 0.001) (Fig. 8c).

**Discussion**

SRD5A3 is a membrane protein, which is mainly located in cellular microsomes and nuclei, and belongs to the steroid 5α reductase family. As a microsomal enzyme, it participates in the reduction reaction of steroid compounds by reducing hydrogen provided by NADPH. Studies have shown that it can promote the occurrence of prostatic hyperplasia and prostate cancer by regulating androgen metabolism. Besides, SRD5A3 reduces polyprenols to polyterpenoids and participates in the synthesis of N-glycosylated oligosaccharide precursors, leading to congenital glycosylation disorders in patients [13,14]. What's more, SRD5A3 is involved in the occurrence of various malignant biological behaviors of human tumors, and is a key gene in the development of tumors. The role of SRD5A3 in a variety of malignancies has been widely reported [15]. However, there are few reports on the role of SRD5A3 in liver cancer.

In this study, bioinformatics analysis using high throughput RNA-sequencing data from TCGA demonstrated that SRD5A3 was significantly highly expressed in HCC tissues compared with paired normal tissues, indicating that SRD5A3 played a role in tumorigenesis and progression. In addition, the ROC analysis showed that the AUC was 0.849 in the diagnosis of HCC, suggesting that SRD5A3 might be a potential diagnostic biomarker. On this premise, the connection between the statement of SRD5A3 and clinicopathological boundaries was additionally examined, and it was discovered that high articulation levels of SRD5A3 protein expression was significantly associated with T stage (p < 0.001), pathologic stage (p < 0.001), tumor status (p = 0.008), race (p = 0.001) and residual tumor (p = 0.017). Despite numerous reports on the role of the SRD5A3 in the development of HCC, the prognostic value of SRD5A3 has rarely been investigated. We constructed a prognostic gene signature model based on SRD5A3 Kaplan–Meier curve indicated a good performance of this model for HCC survival prediction. We found that HCC patients in the high SRD5A3 expression group had poorer OS, PFI and DSS. Multivariate analysis showed that SRD5A3 was an independent factor affecting survival of HCC patients (P<0.05), which proves that it may become a biomarker of HCC.

A significant conclusion in our study concerns the SRD5A3 expression correlation with immune infiltration levels in HCC. Numerous studies have demonstrated that the hepatocellular carcinoma microenvironment contributes to the immunological changes during the progression of HCC. This finding suggested that SRD5A3 might play an important role in the immune system. Thus, the difference in immune cell infiltration between patients with high and low SRD5A3 expression was compared in this study. SRD5A3 was negatively Th17 (R = -0.238, p < 0.001), Cytotoxic cells (R = -0.234, p < 0.001), whereas it was positively correlated with Th2 cells (R = 0.258, p < 0.001). Tumor microenvironment, as an important structural function that mainly supports and promotes the occurrence, development, recurrence and metastasis of tumors, has been paid attention and recognized by many scholars in recent years [16,17]. In the context of precision medicine, a comprehensive analysis of tumor infiltrating cells, cytokines/chemokines, gene expression and other aspects to develop personalized precision
immunotherapy for liver cancer patients is of great significance for the effective evaluation and prediction of the efficacy of immunotherapy and the adoption of combined therapy strategy.

Another important aspect of this study revealed the potential value of SRD5A3 in the progression of HCC cell cycle. To further explore the key pathways in HCC, we performed GSEA using TCGA data, GSEA showed that cell cycle mitotic, cell cycle checkpoints, mitotic nuclear division, o-glycan processing, mitotic rometaphase protein o-linked glycosylation were differentially enriched in SRD5A3 high expression phenotype. Tumor was considered as a cell cycle disease. Researches reveal that abnormal cell cycle, cell apoptosis, and DNA repair play an important role in drug resistance of tumors. For example, abnormal cell cycle signal transduction pathway can restore effective circulation disorder of drug-resistant tumor cells and return to mitosis, leading to resistance to chemotherapy drugs. With long-term use, chemotherapeutic drugs, such as sorafenib, have additional issues such as toxicity and/or drug inefficacy [18]. Chemotherapy resistance has become one of the major obstacles to cancer remission.

In recent years, it has become a hot topic to study the mechanism of chemotherapeutic drug resistance from the perspective of genetics to guide clinical treatment practice. Recent studies have shown that small stress molecules such as arginine and anion small stress molecules acetylcholine chloride have the potential to regulate p53 mutation polymerization [19]. Robert J. Schleifer et al. suggested that decreased KLHL5 expression increased sensitivity to anticancer compounds, especially Akt/PI3K/mTOR inhibitors, which implies that inhibiting or regulating KLHL5 could be a potential method of sensitizing cancer cells to specific drugs [20]. The therapeutic targeting of cyclin-dependent kinase inhibitors (CKI) with anticancer drugs or related genes increases chemotherapy resistance by reducing cyclin E levels or activity in tumor cells [21]. We tried to predict the potential connection between SRD5A3 and the cell cycle, and further provided a new idea for the exploration of chemotherapy drug resistance and the discovery of new biomarker therapeutic target. At the same time, the identification of molecular markers and the classification of HCC patients with high heterogeneity to determine whether they are suitable for specific (targeted or untargeted) therapy will play an important role in guiding the individualized precise treatment of HCC.

There were some limitations to our study that need to be recognized. First, this is a preliminary study designed to investigate the role played by SRD5A3 in HCC. Further experimental verifications are necessary to elucidate the biological functions of these predicted molecular mechanisms in HCC to deepen our understanding of the direct impact of SRD5A3 on HCC. Besides, the sample size in our study was relatively small, and larger sample sizes are needed to increase the reliability of our findings. Last, the clinical information from the TCGA database was not comprehensive, and more clinical data concerning tumor progression and prognosis should be included to better evaluate the relationship between the SRD5A3 and HCC.

Conclusion
In this study, we conducted a comprehensive bioinformatics analysis and found that high expression of SRD5A3 was significantly associated with poor prognosis in HCC patients. It may be a potential biomarker for HCC. The important role of SRD5A3 in tumor immune microenvironment and chemotherapeutic drug resistance was further explored through immune infiltration analysis and GSEA enrichment analysis, so as to provide research ideas for precise and individualized treatment of liver cancer. However, the limitations of this study lie in the small sample size and lack of experimental verification. Therefore, more population-based studies with larger sample sizes and further functional studies are still necessary to confirm our findings.

**Abbreviations**

ACC: adrenocortical carcinoma; BLCA: bladder urothelial carcinoma; BRCA: breast carcinoma, CESC: Squamous cell and adenocarcinoma of the cervix; CHOL: cholangiocarcinoma; COAD: colorectal cancer; DLBC: diffuse large B cell lymphoma; ESCA: esophagus cancer; GBM: Glioblastoma Multiforme; KICH: chromophobe cell tumor; KIRC: clear cell carcinoma of kidney; KIPP: Renal papillary cell carcinoma; LAML: acute myeloid leukemia; LIHC: hepatocellular carcinoma; LGG: Low-grade gliomas; LUAD: lung adenocarcinoma; LUSC: human lung squamous carcinoma; OV: ovarian serous cystadenocarcinoma; PAAD: pancreatic carcinoma; PRAD: prostatic carcinoma; READ: rectum adenocarcinoma; SARC: sarcoma; SKCM: Cutaneous Melanoma; STAD: gastric carcinoma; TGCT: carcinoma of testis; THCA: Thyroid Carcinoma; THYM: thymic carcinoma; UCEC: endometrial carcinoma of uterus; UCS: sarcoma of uterus; AUC: Area under the curve; GO: Gene ontology; GO-BP: GO biological process; GO-CC: GO cellular component; GO-MF: GO molecular function; KEGG: Kyoto Encyclopedia of Genes and Genomes; MCODE = The Molecular Complex Detection; RT-qPCR: Real-time quantitative polymerase chain reaction;

**Declarations**

**Acknowledgments**

The results shown here are in whole or part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.

**Authors’ contributions**

This research was conducted in collaboration with all authors. Jing Xue performed the data curation, analysis, and wrote the first draft of the manuscript. Xianzhao Yang and Feng Jiang reviewed previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding**
This work has been financially supported by the second batch of scientific research projects for the construction of national TCM clinical research base (JDZX2015181).

Availability of data and materials

The raw data of this study are derived from the TCGA (https://portal.gdc.cancer.gov/) and GTEx (http://commonfund.nih.gov/GTEx/), which are publicly available databases.

Ethics approval and consent to participate

Not necessary.

Consent for publication

Not applicable.

Conflict of interest statement

The authors declare that there are no conflicts of interest.

References

1. Chen Z, Xie H, Hu M, Huang T, Hu Y, Sang N, Zhao Y. Recent progress in treatment of hepatocellular carcinoma. Am J Cancer Res. 2020;10(9):2993-3036.
2. Grandhi MS, Kim AK, Ronnekleiv-Kelly SM, Kamel IR, Ghasebeh MA, Pawlik TM. Hepatocellular carcinoma: From diagnosis to treatment. Surg Oncol. 2016;25(2):74-85.
3. Anwanwan D, Singh SK, Singh S, Saikam V, Singh R. Challenges in liver cancer and possible treatment approaches. Biochim Biophys Acta Rev Cancer. 2020;1873(1):188314.
4. Wei R, Zhong S, Qiao L, Guo M, Shao M, Wang S, Jiang B, Yang Y, Gu C. Steroid 5α-Reductase Type I Induces Cell Viability and Migration via Nuclear Factor-κB/Vascular Endothelial Growth Factor Signaling Pathway in Colorectal Cancer. Front Oncol. 2020;10: 1501.
5. Reichardt JK, Makridakis N, Henderson BE, Yu MC, Pike MC, Ross RK. Genetic variability of the human SRD5A2 gene: implications for prostate cancer risk. Cancer Res. 1995;55(18):3973-5.
6. Uemura M, Tamura K, Chung S, Honma S, Okuyama A, Nakamura Y, Nakagawa H. Novel 5 alpha-steroid reductase (SRD5A3, type-3) is overexpressed in hormone-refractory prostate cancer. Cancer Sci. 2008;99(1):81-6.
7. Park JM, Song KH, Lim JS, Kim JW, Sul CK. Is the Expression of Androgen Receptor Protein Associated With the Length of AC Repeats in the Type III 5-α Reductase Gene in Prostate Cancer
Patients? Korean J Urol. 2013;54(6):404-8.

8. Uemura M, Tamura K, Chung S, Honma S, Okuyama A, Nakamura Y, Nakagawa H. Novel 5 alpha-steroid reductase (SRD5A3, type-3) is overexpressed in hormone-refractory prostate cancer. Cancer Sci. 2008;99(1):81-6.

9. Wang ZH, Zhang YZ, Wang YS, Ma XX. Identification of novel cell glycolysis related gene signature predicting survival in patients with endometrial cancer. Cancer Cell Int. 2019;19: 296.

10. Qi F, Qin WX, Zang YS. Molecular mechanism of triple-negative breast cancer-associated BRCA1 and the identification of signaling pathways. Oncol Lett. 2019;17(3):2905-2914.

11. Martin B, Paesmans M, Berghmans T, Branle F, Ghisdal L, Mascaux C, Meert AP, Steels E, Vallot F, Verdebout JM, Lafitte JJ, Sculier JP. Role of Bcl-2 as a prognostic factor for survival in lung cancer: a systematic review of the literature with meta-analysis. Br J Cancer. 2003;89(1):55-64.

12. Jochems C, Schlom J. Tumor-infiltrating immune cells and prognosis: the potential link between conventional cancer therapy and immunity. Exp Biol Med (Maywood). 2011;236(5):567-79.

13. Stiles AR, Russell DW. SRD5A3: A surprising role in glycosylation. Cell. 2010;142(2):196-8.

14. Jaeken J, Lefeber DJ, Matthijs G. SRD5A3 defective congenital disorder of glycosylation: clinical utility gene card. Eur J Hum Genet. 2020;28(9):1297-1300.

15. Li J, Ding Z, Wang Z, Lu JF, Maity SN, Navone NM, Logothetis CJ, Mills GB, Kim J. Androgen regulation of 5α-reductase isoenzymes in prostate cancer: implications for prostate cancer prevention. PLoS One. 2011;6(12): e28840.

16. Hui L, Chen Y. Tumor microenvironment: Sanctuary of the devil. Cancer Lett. 2015;368(1):7-13.

17. Petitprez F, Meylan M, de Reyniès A, Sautès-Fridman C, Fridman WH. The Tumor Microenvironment in the Response to Immune Checkpoint Blockade Therapies. Front Immunol. 2020;11: 784.

18. Anwanwan D, Singh SK, Singh S, Saikam V, Singh R. Challenges in liver cancer and possible treatment approaches. Biochim Biophys Acta Rev Cancer. 2020;1873(1):188314.

19. Chen Z, Kanapathipillai M. Inhibition of p53 Mutant Peptide Aggregation In Vitro by Cationic Osmolyte Acetylcholine Chloride. Protein Pept Lett. 2017;24(4):353-357.

20. Schleifer RJ, Li S, Nechtman W, Miller E, Bai S, Sharma A, She JX. KLHL5 knockdown increases cellular sensitivity to anticancer drugs. Oncotarget. 2018;9(100):37429-37438.

21. Pang W, Li Y, Guo W, Shen H. Cyclin E: a potential treatment target to reverse cancer chemoresistance by regulating the cell cycle. Am J Transl Res. 2020;12(9):5170-5187.

22. Vivian J, Rao A A, Nothaft F A, et al. Toil enables reproducible, open source, big biomedical data analyses[J]. Nature biotechnology.2017; 35(4): 314-316.

23. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research .2015;43(7): e47.

24. Subramanian A, Tamayo P, Mootha V K, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles[J]. Proceedings of the National Academy of Sciences.2005;102(43): 15545-15550.
25. Ref: Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC bioinformatics. 2013;14:7.

26. Bindea, Gabriela, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity. 2013;39(4):782-95.

**Figures**

Figure 1

Differential expression map of SRD5A3. (a) Differential expression of SRD5A3 in tumor and normal tissues in TCGA and GTEx pan-cancer data. (b) the expression of SRD5A in GTEX combined with TCGA normal samples and TCGA hepatocellular carcinoma samples (c) SRD5A3 expression in normal and tumor tissues in HCC from TCGA. (d) Expression of SRD5A3 in HCC cancer tissues and corresponding normal tissues in the TCGA database. (e) ROC curves of SRD5A3 expression to predict HCC.
Figure 2

Differential expression genetic map of HCC in the TCGA database. (a) It is a volcano plot. The x-axis is the logarithm of $-\log_{10}$ after correction. The larger the value, the more significant the difference. The y-axis is log2(FC), and the larger the absolute value, the greater the fold change. (b) It is a heatmap. Blue represents high expression, and yellow represents a low expression.
Figure 3

Associations between the SRD5A3 expression and Gender(a), Race(b), T stage(c), pathologic stage(d), tumor status(e), residual tumor(f).
Figure 4

Survival analysis of SRDA3 in HCC. Overall survival curve of patients with high versus low SRD5A3 expression levels.
Figure 5

(a) Nomogram for predicting the probability of 1-, 3-, and 5-year OS for HCC patients (b) Calibration plot of the nomogram for predicting the probability of OS at 1, 3, and 5 years.

Figure 6

The enrichment analysis of SRD5A3 and neighboring genes in OC (Metascape) (a) Heatmap of Gene Ontology (GO) enriched terms colored by p-values. (b) Network of GO enriched terms colored by p-value, where terms containing more genes tend to have a more significant p-value. (c) Heatmap of Kyoto Encyclopedia of Genes and Genomes (KEGG) enriched terms colored by p-values. (d) Protein–protein interaction (PPI) network (e) five most significant MCODE components form the PPI network. (f) Independent functional enrichment analysis of four MCODE components.
Figure 7

Enrichment plots from gene set enrichment analysis (GSEA). GSEA results showed that cell cycle mitotic(a), cell cycle checkpoints (b), mitotic nuclear division (c), O-glycan processing(d), mitotic prometaphase (e), protein O-linked glycosylation (f) were enriched mainly in SRD5A3-related HCC.
Figure 8

Relationship between SRD5A3 expression and tumor-infiltrating immune cells, including Th17 cell (a), cytotoxic cell (b), and Th2 cell (c).