Aus der
Klinik für Allgemein-, Viszeral- und Transplantationschirurgie
Klinikum der Ludwig-Maximilians-Universität München

Clinical value of Plasmalemma Vesicle Associated Protein in Hepatocellular Carcinoma

Dissertation
zum Erwerb des Doktorgrades der Medizin
an der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

vorgelegt von
Zhongyi Li
aus
Hubei
Jahr
2025

Mit Genehmigung der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

Erstes Gutachten:	Prof. Dr. Jens Werner
Zweites Gutachten:	Prof. Dr. Christian Lange
Drittes Gutachten:	Priv. Doz. Dr. Stefan Gölder
Weitere Gutachten:	
Mitbeurteilung durch die promovierten Mitarbeiter:	Dr. Frank K.Braun und Dr. Dominik Koch
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	16.06.2025

Zusammenfassung (Deutsch)

Hintergrund: Das Hepatozelluläre Karzinom (HCC) ist der häufigste primäre Lebertumor und gehört zu den häufigsten bösartigen Tumoren weltweit. Aufgrund der hohen Morbidität und Mortalität stellt es eine erhebliche Herausforderung für das Gesundheitswesen dar. Hierzu trägt bei, dass das HCC im Frühstadium keine offensichtlichen Symptome aufweist und die optimale Behandlungszeit oft verpasst wird. Aufgrund der trotz verschiedenster mittlerweile zur Verfügung stehender Therapiemodalitäten schlechten Prognose besteht dringender Bedarf an einem noch tieferen Verständnis des HCC um die Therapieoptionen weiter zu verbessern.

Methoden: Mehrere HCC-Datensätze wurden für eine differenzielle Analyse verwendet. Zwei Methoden des maschinellen Lernens, Lasso und SVM-RFE, wurden angewendet um Schlüsselgene zu identifizieren, die differenziell exprimiert sind. Gleichzeitig wurde auf diagnostische und prognostische Aussagekraft dieser Gene untersucht. Einzelzellsequenzierungsmethoden wurden verwendet, um die Expression von Genen in der Tumormikroumgebung zu analysieren. Es erfolgte eine Korrelationsanalyse, um die Beziehung zwischen Genen und der Expression allgemeiner Biomarker zu verstehen. In einem zweiten Schritt verwendeten wir klinische HCC-Gewebeproben und führten eine Hämatoxylin- Eosin-Färbung sowie eine Immunfluoreszenzfärbung durch, um die Expressionsorte und Expressionsniveaus von identifizierten Genen in Geweben zu untersuchen. Schließlich kombinierten wir Genexpression und klinische Indikatoren, um den klinischen Wert dieser Gene beurteilen zu können.

Ergebnisse: Es wurden 7 Gene sowohl über Lasso, als auch SVM-RFE identifiziert. Alle 7 Gene zeigten in mehreren HCC-Datensätzen eine gute diagnostische Aussagekraft. Die weitere Analyse dieser 7 Gene ergab, dass PLVAP eine prognostische Aussagekraft zum HCC-Gesamtüberleben (OS) aufweist. Die Ergebnisse der Einzelzellsequenzierung legen nahe, dass PLVAP in Endothelzellen in HCC-Geweben stark exprimiert wird. Ergebnisse der Immunfluoreszenz bestätigten diese Ergebnisse der Einzelzellsequenzierung. PLVAP zeigte eine moderate

Korrelation mit den endothelialen Biomarkern CD31, CD34, CD105 und VEGFR2. Wir haben herausgefunden, dass der klinische Wert von PLVAP darin besteht, dass eine Korrelation zwischen PLVAP und der Tumorgröße (T-Stadium) des Patienten besteht. Patienten mit hoher PLVAP-Expression haben ein niedriges T-Stadium, während Patienten mit niedriger PLVAP-Expression eine hohes T-Stadium aufweisen.

Schlussfolgerung: PLVAP korreliert mit klinischen Daten bei HCC Patienten. Einzelzellsequenzierungsdaten und Immunfluoreszenzfärbung bestätigten, dass PLVAP hauptsächlich in vaskulären Endothelzellen vorhanden ist und die PLVAP-Expression mit dem T-Stadium von HCC Patienten korreliert.

Abstract (English)

Background: One of the most common malignant tumors in the world is hepatocellular carcinoma (HCC). The high mortality and morbidity pose a serious threat to human health. Although various treatment methods are available, symptoms of HCC at the early stage are not obvious, and the best treatment time has often been lost after the diagnosis. For further improvement of diagnosis and therapy, it is urgent to have a deeper understanding of HCC.

Methods: Multiple HCC datasets were combined for differential analysis of HCC. Differentially expressed genes screened for key genes in HCC were performed using Lasso and SVM-RFE machine learning methods. At the same time, our study also examined key genes diagnostic and prognostic capabilities. We also used single-cell sequencing data to analyze and clarify the expression levels of genes in the tumor microenvironment. Moreover, correlation analysis was used to determine the relationship between gene expression and common biomarkers. Finally, we used HCC tissue specimens from the clinic, performed Hematoxylin-Eosin staining and immunofluorescence staining, and explored the expression site and expression level of genes in tissues. Finally, we combined gene expression levels with clinical indicators to explore the clinical value of the genes.

Results: A set of 7 genes were identified by the two independent analysis methods, Lasso regression and SVM-RFE. All 7 genes showed good diagnostic value in multiple HCC data sets. We found that PLVAP showed prognostic value in HCC overall survival (OS). The results of single-cell sequencing data suggested that PLVAP is highly expressed in endothelial cells in HCC tissues. The results from immunofluorescence also confirmed the results of single-cell sequencing data. PLVAP showed moderate correlation with vascular endothelial biomarkers CD31, CD34, CD105, and VEGFR2. Analyzing clinical data, we found that PLVAP is correlated with patients T stage. Patients with high expression of PLVAP have a lower T stage, and conversely, patients with low expression of PLVAP have higher T stage.

Conclusion: PLVAP correlates with clinical data in HCC. Subsequent single-cell sequencing data and immunofluorescence staining confirmed that PLVAP is mainly present in vascular endothelial cells, and PLVAP expression is correlated with the patients T stage.

List of Figures

The colors red and blue indicate high expression and low expression, respectively.

Figure 5. Single-cell sequencing analysis of PLVAP expression comes from GSE166635. A. Expression of PLVAP in tumor cells, stromal cells, and immune cells. B. The expression of PLVAP in various types of tumor cells, stromal cells, and immune cells. Depending on the depth of the red, the greater the expression.

Figure 6. Analysis of the correlation between the expression of PLVAP and endothelial cell biomarkers in HCC. A. Correlation analysis between PLVAP and CD31 (PECAM1), correlation coefficient R=0.57. B. Correlation analysis between PLVAP and CD34, correlation coefficient R=0.69. C. Correlation analysis between PLVAP and CD105 (ENG), correlation coefficient R=0.57. D. Correlation analysis between PLVAP and VEGFR2 (KDR), correlation coefficient R=0.59.

Figure 7. Three HCC patients' tumor tissues and adjacent tissues stained with HE staining and IF staining for PLVAP (AN16969, AN40667 and AN24563). A. Tumor tissue and adjacent tissues of a patient with HCC (AN16969) stained with HE. B. Tumor tissue and adjacent tissues stained with PLVAP immunofluorescence in patient AN16969 with HCC. C. The tumor tissue and adjacent tissues of HCC patient AN40667 were stained with HE. D. Tumor tissue and adjacent tissue of HCC patient AN40667 stained with PLVAP immunofluorescence. E. The tumor tissue and adjacent tissue from patient AN24563 were stained with HE. F. The tumor and adjacent tissues of HCC patient AN24563 were stained with PLVAP immunofluorescence.

Figure 8. Summary of PLVAP expression in HCC patient samples. A. The patients were ranked from low to high according to the expression level of PLVAP in their tumor tissues. B. Comparison of PLVAP expression in tumor and adjacent tissues in 19 pairs of HCC patients. C. Each patient's level of PLVAP expression in adjacent tissues.

Figure 9. Correlation analysis between PLVAP expression and T stage in HCC patients. Each HCC patient's PLVAP expression level is represented in blue, and the T stage is represented in orange.

List of Tables

List of abbreviations

HCC	hepatocellular carcinoma
MRI	Magnetic Resonance Imaging
HCV	hepatitis virus C
CT	Computed Tomography
HBV	hepatitis virus B
AFP	alpha-fetoprotein
SBRT	Stereotactic body radiation therapy
PDGFR	platelet-derived growth factor receptor
EGFR	epidermal growth factor receptor
VEGFR	vascular endothelial growth factor receptor
INSR	insulin receptor
TACE	trans arterial chemoembolization
OS	overall survival
PFS	progression-free survival
Lasso	Least Absolute Shrinkage and Selection Operator
SVM-RFE	Support Vector Machine Recursive Feature Elimination
AMLWVF	angiomyolipoma without visible fat
RCC	renal cell carcinoma
TME	tumor microenvironment
PLVAP	Plasmalemma Vesicle Associated Protein
HCl	Hydrochloric acid
NaCl	Sodium chloride
BSA	Bovine Serum Albumin
GEO	Gene Expression Omnibus
NCBI	National Center for Biotechnology Information
NCI	National Cancer and Oncology Institute
NHGRI	National Human Genome Research Institute
TCGA	The Cancer Genome Atlas
GTEx	Genotype-Tissue Expression
ROC	Receiver Operating Characteristic
DEGs	differentially expressed genes
NIH	National Institutes of Health
AUC	Area Under the Curve
GEPIA	Gene Expression Profiling Interactive Analysis
KM	Kaplan-Meier
CD31	Cluster of Differentiation 31
PECAM1	Platelet Endothelial Cell Adhesion Molecule-1
ENG	Endoglin

VEGFR2	Vascular Endothelial Growth Factor Receptor 2
RMSE	root mean square error
λ	lambda
CLEC4M	C-Type Lectin Domain Family 4 Member M
ECM1	Extracellular Matrix Protein 1
DBH	Dopamine Beta-Hydroxylase
CFP	Complement Factor Properdin
NAT2	N-Acetyltransferase 2
CXCL14	C-X-C Motif Chemokine Ligand 14
Treg cells	regulatory T cells
Tprolif cells	proliferating T cells
DC cells	Dendritic cells
TISCH	Tumor Immune Single Cell Hub
Mast.	Mast Cells
Endoth.	Endothelial cells
Fibro.	Fibroblasts
Epithel.	Epithelial cells
Malig.	Malignant cells
NAFLD	Nonalcoholic fatty liver disease
Mono./Macro.	Monocytes/Macrophages

1. Introduction

1.1 Hepatocellular Carcinoma

1.1.1 Epidemiology of Hepatocellular Carcinoma

Primary liver cancer is one of the most common tumor types in the world. As a result of its high morbidity and mortality, it has a serious adverse effect on the health of the global population. Liver cancer ranks sixth in terms of tumor incidence, but third in cancer-related deaths worldwide [1, 2]. Liver cancers can be classified as hepatocellular carcinomas (HCC) and intrahepatic cholangiocarcinoma. Hepatocellular carcinoma is the most frequent subtype with 80% of all liver cancer diagnoses [3]. According to the World Health Organization's latest data (Cancer Today (iarc.fr)) on mortality of liver cancer worldwide, Asia accounts for approximately 70%, and Europe for approximately 10% of liver cancer incidences (Figure 1). Despite accounting for 21% of the global population, Asia has 54.3% of liver cancer incidence and 54.1% of liver cancer mortality. Additionally, men have a higher incidence and mortality rate of liver cancer than women [4]. It is estimated that in 2040, the number of people diagnosed with liver cancer will be about 1.4 million, and approximately 1.3 million will die from liver cancer [5]. China has a significantly higher incidence and mortality rate for liver cancer than anywhere else in the world, accounting for 45.3% of all liver cancer cases [5]. In contrast to China, states such as South Korea and Japan report a much smaller mortality for liver cancer, despite a similar high incidence rate of liver cancer. This is in part attributed to extensive early screening for liver cancer [6, 7]. There has been a significant change in the epidemiology of liver cancer in Europe in the past few decades. While liver cancer incidence is lower in Europe than in Asia, this has changed significantly over the past few decades. In addition to Europe’s aging population structure, obesity-associated metabolic-related non-alcoholic steatohepatitis and diabetes contribute to the rapid increase in liver cancer diagnosis. Metabolic-related non-alcoholic steatohepatitis can develop into cirrhosis and

eventually into liver cancer [8-10]. Similar to South Korea and Japan, the increase in liver cancer incidences does not correlate in Europe with an increase in mortality. This has been attributed to the presence of advanced health care systems [11]. Most European countries provide cancer screening programs enabling early diagnosis with subsequent treatments. In combination, these efforts effectively reduced the mortality rate of liver cancer.

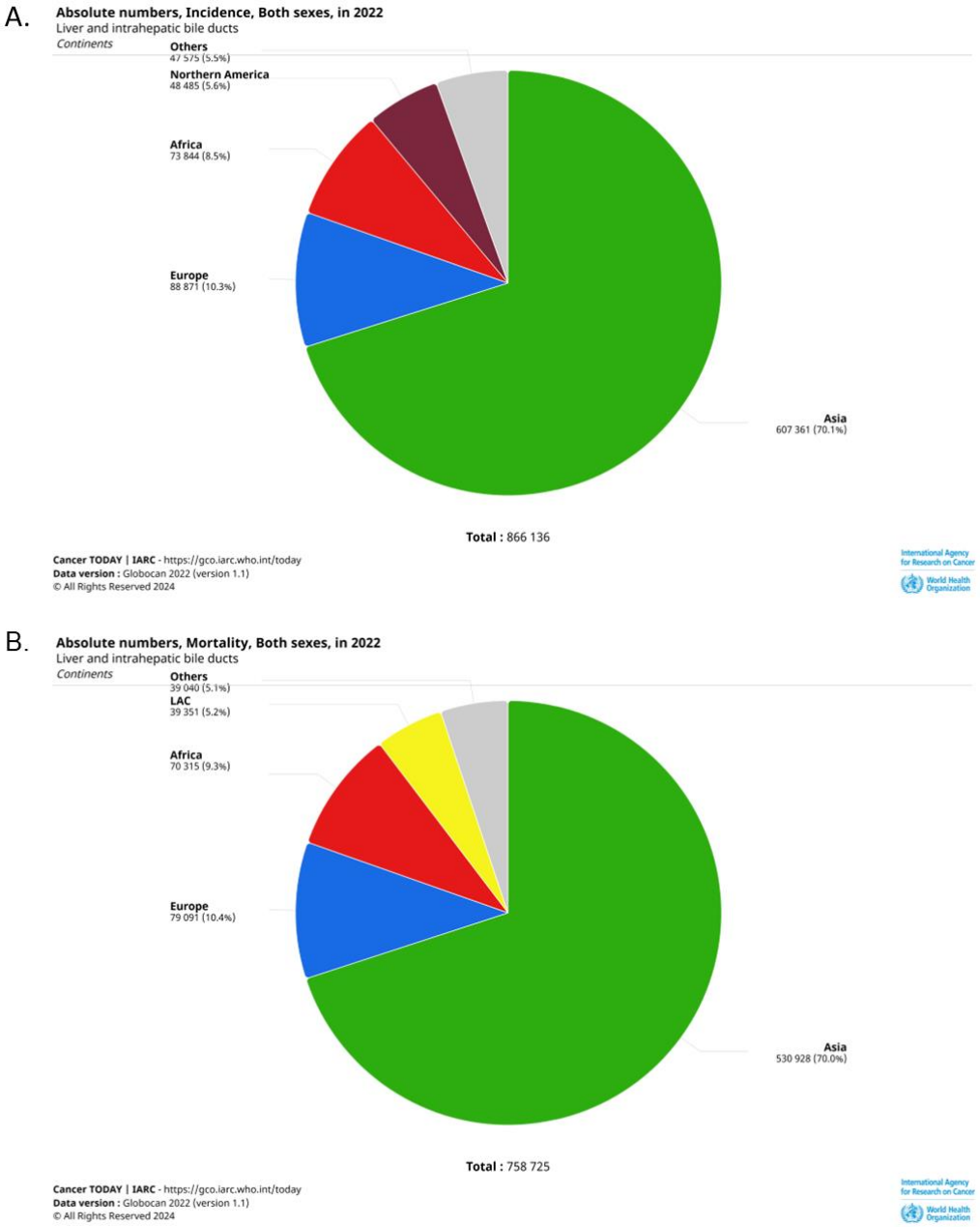

Figure 1. Incidence and mortality of liver cancer in different regions in 2022, according to the World Health Organization (Cancer Today (iarc.fr)). A. The incidence of liver cancer in different regions of the world. B. Liver cancer mortality in different regions of the world.

Medical research has shown that a viral infection such as hepatitis B virus (HBV) and hepatitis C virus (HCV) may also contribute to HCC incidence [12, 13]. Other risk factors include alcohol-related liver disease, non-alcoholic steatohepatitis, and high intake of aflatoxin [14]. However, HBV and HCV are recognized as major causes for HCC in the world, accounting for roughly two-thirds of all HCC diagnosis [15]. With respect to the impact of viral infections, significant differences can be found between Eastern and Western populations. HBV is described as a major risk factor in Asia and Africa, accounting for 60% of the total risk factors, and only 20% for the Western world. Meanwhile, HCV is a major risk factor in Western countries, where it can reach up to 44% in the total risk factors [16]. Due to an active and widespread anti-hepatitis virus vaccination effort, studies found a slow decline in virus-related HCC cases, but incidence of metabolic liver disease-related and alcohol-related HCC continue to increased significantly [17, 18].

In addition to the diseases mentioned above that can cause liver dysfunction and liver cancer, liver disease caused by nonalcoholic steatohepatitis (NASH) has attracted increasing attention in recent decades. Over the past few decades, nonalcoholic steatohepatitis has become more common with the increasing obesity and diabetes rates. This is a particular problem for Western societies [19, 20]. The pathological characteristics of NASH are fat accumulation in liver tissues accompanied by inflammation, cellular damage, death of liver cells, and fibrosis [20]. Although not all patients with nonalcoholic steatohepatitis develop liver cancer, epidemiological studies correlate an increase in incidences of liver cancer with increased diagnosis of NASH, particularly in developed countries [21], which resulted in NASH becoming a significant risk factor that has to be addressed. In addition to affecting the liver, NASH is shown also to affect other organs, further elevating the need for additional medical research and preventative programs.

In conclusion, as countries around the world implement extensive HBV vaccination programs and continue with the research on antiviral treatment options, virus-related liver cancer incidences will continue to dwindle. However, lifestyle-related obesity and

diabetes with subsequent development of NASH will fuel new incidences of liver cancer. A redirecting of research efforts is needed to address these new causes of liver cancer.

1.1.2 Diagnosis of HCC

With the tremendous progress in medicine in recent decades, the diagnosis of HCC has also been significantly improved. Existing diagnostic methods for HCC include imaging examination, serum marker detection, histopathological examination, molecular diagnosis, etc. Imaging examinations mainly include Computed Tomography (CT), Nuclear Magnetic Resonance Imaging (MRI), and ultrasound examination. Among them, abdominal ultrasound examination is considered the first choice for HCC because of its non-invasiveness, simplicity and low cost [22]. The examination of serum markers mainly focuses on alpha-fetoprotein (AFP). Most HCC patients have abnormally high levels of serum AFP. However, high serum AFP alone does not necessarily mean the patient has HCC [23]. Ovarian endodermal sinus tumor and choriocarcinoma have been shown to have increased AFP levels [24]. For a final HCC diagnosis additional parameter need to be checked. So far, the diagnostic methods mentioned above are helpful for the diagnosis of HCC, but histopathological biopsy are still essential to confirm an HCC diagnosis [25]. Histopathological examination of liver tissues obtained from puncture biopsy or after surgical resection is considered the only gold standard for diagnosing HCC [26]. Pathologists observe atypia under a microscope. Atypia is the morphological difference between abnormal cells and normal cells. It mainly includes cellular atypia and structural atypia. Cellular atypia includes the size and shape of cells, and whether the nuclear division is frequent. Tumor cells generally increase in size, the nucleus increases, the proportion of the nucleus to the entire cell also increases, and the nuclear division is relatively frequent. Structural atypia is the difference between the spatial structural arrangement of the diseased tissue and normal tissue. Under a microscope, it can be observed that the tumor tissue is arranged in a disordered manner [27]. In addition to observing morphological changes under a microscope, some common markers are also helpful for tumor

identification. Through immunohistochemical staining, the marker of Proliferation, Ki-67(MKI67), can represent the high proliferation characteristics of tumor cells [28]. Germ cell tumors specifically express Alkaline Phosphatase (PLAP) [29], while prostate cancer express Kallikrein Related Peptidase 3 (KLK3) [30]. Molecular diagnosis of HCC includes gene mutation analysis, DNA methylation analysis, and liquid biopsy analysis. The most frequently mutated genes in HCC are Telomerase Reverse Transcriptase (TERT), Catenin Beta 1 (CTNNB1), and Tumor Protein P53 (TP53) [31]. Mutational status of these genes can be used to diagnose early liver cancer and predict the prognosis of patients with HCC [32]. During DNA methylation, methyl groups are added to cytosine, which does not change the DNA sequence but can impact a cell's ability to access methylated DNA for the transcription of genes. It is well-established that the pathogenesis of HCC is influenced by DNA methylation [33]. Early stages of liver cancer can be detected using DNA methylation detection with good specificity [34]. Liquid biopsy is different from traditional tissue biopsy. Liquid biopsies are conducted by collecting peripheral blood from patients and testing it for circulating tumor DNA (ctDNA) and circulating tumor cells. Liquid biopsy is less invasive than traditional tissue biopsies and with the tremendous progress in the past two decades in improving the sensitivity in detecting small amounts of DNA and the isolation of a small number of tumor cells, liquid biopsies have become a staple in many cancer screening techniques [35].

As mentioned above, there are many diagnostic methods for HCC, though each method comes with benefits and limitations. Early detection of HCC would provide patients with better treatment options; however, early stages of liver cancer are usually without obvious symptoms and thus go unnoticed [36]. AFP, in the clinical practice commonly used HCC marker, lacks sufficient sensitivity and specificity and is therefore not reliable for diagnosis [37]. Imaging diagnostic methods such as ultrasound, CT, and MRI often fail in detecting small tumors in the liver. At the same time, benign nodules in the liver, such as hemangiomas, are not easy to distinguish from tumors with these imaging methods alone [38]. However, CT and MRI are quite expensive and

are therefore not suitable as widely used cancer screening tools. Despite liver biopsy being the gold standard for diagnosing liver cancer, it is an invasive procedure that carries a number of risks due to infections, bleeding, and potential tumor cell dissemination [39]. Molecular diagnostic technology has developed rapidly and has shown great potential, but high medical costs limit its widespread use [40].

Like many cancers, HCC is a highly heterogeneous tumor. It is not only possible for tumors to be heterogeneous between them, but also within them [41]. The heterogeneity among HCC patients cannot be reflected and evaluated by a single technology and requires a combination of multiple technologies and methods for evaluation [42].

Therefore, the diagnosis of HCC should be a comprehensive application of imaging, serum markers, biopsies, and molecular diagnosis. The diagnosis of HCC is increasingly becoming individualized, early, and precise with the development of science and technology. How to develop diagnostic markers with high sensitivity, high specificity, reasonable prices and minimal trauma to the human body is the future research direction for HCC diagnosis, which requires further in-depth research and extensive clinical trials.

1.1.3 Treatment of HCC

The treatment methods for HCC have been improved with the development of science and technology. There are several treatment options available for HCC, including surgical resection, liver transplantation, ablation therapy, radiotherapy, chemotherapy, immunotherapy, and molecular targeted therapy.

For HCC patients with good liver function (Child-Pugh A), a small number of tumor lesions in the liver, no vascular invasion, and no lymph node metastasis, surgical resection is the preferred treatment approach [43]. However, 70% of the patients will have a relapse within five years after surgery [44].

The Milan criteria for liver transplantation require that the diameter of a single cancer lesion should not exceed 5 cm, or the number of multiple cancer lesions should not

exceed 3 with the largest diameter not exceeding 3 cm. The tumor should not invade the liver's large blood vessels or metastasize to distant sites. Patients who meet the Milan criteria can be considered for liver transplantation. However, liver transplantation surgery requires matching donor livers, which are often not available, limiting the treatment approach [45].

Local ablation of HCC is performed under the guidance of ultrasound or CT. The ablation needle is punctured into the tumor and kills tumor cells by releasing energy or chemicals into the tumor tissue, generating local high temperatures or dehydrating the tumor tissue. Like liver cancer resection, liver cancer ablation has a risk of recurrence of up to 70% [46].

Stereotactic body radiation therapy (SBRT) has become more popular in recent years as a form of HCC radiotherapy due to its high focus, accuracy, and low adverse reactions compared with conventional radiotherapy [47]. SBRT focuses radiation from multiple angles on one point, making it a highly precise technique, that can utilize high radiation dosage while still limiting the potential damage to surrounding normal tissue. However, radiotherapy is recommended for patients who do not meet specific criteria to undergo liver surgery but have good liver function, this is the case for around 40% off patients with diagnosed HCC [48, 49].

Traditional chemotherapy is highly toxic and patients suffer from severe side effects, such as liver and kidney dysfunction. Relapse is observed in 53.3% of HCC patients treated with traditional systemic chemotherapy within two years [50]. The results of a meta-analysis showed that the 5-year recurrence rate was approximately 60% in HCC patients who underwent liver resection plus oral chemotherapy [51]. In comparison, trans arterial chemoembolization (TACE) is an improved technique in which a chemotherapy drug cocktail is injected via the hepatic artery. Compared with systemic chemotherapy drug administration, TACE reduces the damage of chemotherapy drugs to other organs which overall increases the quality of life of patients [52].

With the advancements in our understanding of cancer biology, many new potential

druggable signaling pathways have been identified. An improved understanding on their role in different cancer types resulted in the development of specific inhibitors such as sorafenib which inhibits the vascular endothelial growth factor receptor (VEGFR) signaling pathway [53], bevacizumab which inhibits vascular endothelial growth factor (VEGF) [54], and atezolizumab which inhibits programmed cell death 1 ligand 1(PD-L1) [55]. By targeting these signaling pathways a more precise intervention in HCC cancer cells can be achieved, including stopping proliferation as well as induction of apoptosis mediated by immune cells.

Targeting key molecules in important signaling pathways is an important treatment approach. Among the targeting molecules of signaling pathways are growth factor receptors, or molecules the promote angiogenesis [56]. Growth factor receptors are a large family, including insulin receptor (INSR), vascular endothelial growth factor receptor (VEGFR), and epidermal growth factor receptor (EGFR). Studies have found that growth factor receptors can be overexpressed or even mutated in many tumors, promoting cell proliferation and angiogenesis in cancer [57-59]. Sorafenib, a tyrosine kinase inhibitor, mainly acts on signaling pathways such as VEGFR and platelet-derived growth factor receptor (PDGFR) [60]. In many clinical studies sorafenib was shown to effectively prolong the survival time such as overall survival (OS) and progression-free survival (PFS) in HCC patients. Sorafenib became the first "targeted" treatment for HCC to be approved by the US FDA [61, 62]. The main mechanism of action of Ramucirumab, a monoclonal antibody, is targeting VEGFR2. Angiogenesis is inhibited by ramucirumab by blocking VEGF/VEGFR2 signaling, reducing the blood supply to tumor tissue, and ultimately exerting an anti-tumor effect [63]. Pembrolizumab is a PD-1 inhibitor that aims to restore the anti-tumor effect of T cells by blocking the binding of PD-L1 and PD-1 [64]. Clinical trials showed that Pembrolizumab exhibits good anti-tumor effects and effectively prolongs the prognosis time of patients, with respect to OS and PFS [65].

In recent years, immunotherapeutic approaches have made significant progress. With these, the aim is to re-arm or to re-direct the patient’s own immune cells towards tumor

cells. CTLA-4 is a cell surface immune checkpoint that interferes with immune cell function, particularly during the early stages of an antitumor immune response [66, 67]. The CTLA-4 inhibitor ipilimumab binds to CD80 and CD86 to block the anti-immune signaling from cancer cells. With CTLA-4 blocked the immune cells are enabled again to target and kill cancer cells [68].

However, despite all these possible treatment approaches, patients diagnosed with HCC often face the problem of relapse, metastasis, and the development of drug resistance. Most treatment protocols employ multiple drugs and strategies to reduce the risk of relapse and drug resistance. How to solve these problems is of great interest to many physicians and scientists.

1.2 Machine Learning Lasso and SVM-RFE in Medicine

1.2.1 Machine Learning Lasso in Medicine

Least Absolute Shrinkage and Selection Operator (Lasso) is a kind of linear regression model [85]. Lasso is good at processing high-dimensional gene expression matrices. Therefore, in bioinformatics, it is widely used to screen for key genes from thousands of genes based on their characteristics with Lasso [86].

It has been widely used for the assessment of disease biomarkers for early diagnosis and prognosis. For instance, Lasso regression method was used to search for potential biomarkers for Alzheimer's disease, and the results showed that it has good diagnostic value [87, 88]. Core genes were identified using Lasso to screen large-dimensional transcript data sets of HCC, and the possible functions and prognostic abilities of core genes were then explored [89]. This methodology has been previously used to identify prognostic relevant markers such as TOP2A in lung cancer. It has been found that non-small cell lung cancers express high levels of TOP2A and TOP2A was correlated with prognosis by Lasso regression method [90].

1.2.2 Machine Learning SVM-RFE in Medicine

Support Vector Machine Recursive Feature Elimination (SVM-RFE) is one of the machine learning methods that uses the recursive feature elimination mechanism in mathematics to find characteristic genes in large data sets. There are many examples of the use of SVM-RFE in medicine or biomedical research. SVM-RFE was used to find new signature markers in tumors from gene expression matrices [91]. In addition to being used in data mining for high-throughput sequencing, SVM-RFE is also involved in the analysis of MRI data. For this purpose, SVM-RFE is used to extract characteristic MRI images from patients for differential diagnosis [92]. Some researchers have also explored the differentiation of angiomyolipoma without visible fat (AMLWVF) and renal cell carcinoma (RCC). SVM-RFE was used to extract characteristic features from CT images of AMLWVF and RCC to differentiate between AMLWVF and RCC [93].

1.3 Tumor microenvironment

The tumor microenvironment (TME) is in general a term that describes the tissue around the cancer [69]. In addition to the cancer cells, it consists of normal tissue, immune cells (T cells, B cells, NK cells, neutrophils) and other matrix cells (vascular endothelial cells, fibroblasts). TME can vary in the composition and the signaling it provides [70].

By producing growth factors and chemokines, tumor-associated fibroblasts promote the angiogenesis of endothelial cells as well as tumor cell proliferation [71]. Tumor-associated fibroblasts promote tumor cell invasion by degrading proteoglycans in the extracellular matrix [72]. Tumor-associated macrophages promote angiogenesis through angiogenic factors such as VEGFA (Vascular Endothelial Growth Factor A). At the same time, they produce lymphangiogenic factors that promote lymphatic vessel formation in tumor tissues [73].

There are many subtypes of T cells in the immune system. The main groups are CD4

and CD8 T cells, which have critical but distinct roles in the immune system. B cells producing antibodies with which specific structures on the surface of cells can be identified and used as guidance for cytotoxic T cells, for example. Other immune cells can secrete perforin to directly kill bacteria [74, 75]. In the acute phase of an immune response, strong pro-proliferative signaling results in an increased number of cells, which need to be eliminated once the immune response ends. However, not all cells die. Some T cells turn into memory T cells and provide a more rapid response against the same target in the future. An interesting subset of T cells are so call T regulatory cells (Treg cells). Treg cells function as immune suppressors, tasked with preventing excessive immune responses [76]. In many tumors, Treg cells are usually in an activated state, resulting in the suppression of other immune cells. Under these conditions, the immune system cannot effectively eliminate an identified cancer cell. This is called immune escape [77].

B cells exert immune responses by producing antibodies [78], whereas NK cells can directly target cells. NK cells act on target cells through endocytosis and the release of penetrants, or granzymes. Enzymes secreted by these cells cause holes in the cell membranes, which cause the cells to die [79]. NK cells can also secrete various cytokines to directly act on specific signaling pathways in target cells and the surrounding area [80].

Endothelial cells of blood vessels are single-layer flat epithelial cells [81]. The function of vascular cell tissue is to form vascular structures and transport blood and nutrients to tissue [82]. In many TME the cytokines secreted by various cells promote angiogenesis. An increase in blood vessels has the obvious benefit of more oxidation of the area as well as an increase in the nutrition. Therefore, anti-angiogenesis treatments have been devised in recent years and decades. Several drugs targeting the formation of blood vessels in tumors, such as bevacizumab, sorafenib, and ramucirumab, are now available [83]. However, the lack of tissue specificity as well as the complexity of the angiogenesis signaling network and resulting drug resistances have reduced the effectiveness of this approach [84].

1.4 Plasmalemma Vesicle Associated Protein（PLVAP）

Plasmalemma Vesicle Associated Protein (PLVAP), or PV1, is expressed on the surface of cells. PLVAP is a key component of the membranes of caveolae, transendothelial channels, and stomatal diaphragms of vesicle-vacuolar organelles, as well as the diaphragms of endothelial fenestrae [94]. Several studies on PLVAP have identified a close relationship between PLVAP expression and prognosis of gastric adenocarcinoma. The 16S rRNA gene sequencing technology of 15 formalin-fixed, paraffin-embedded gastric cancer specimens was used to analyze the abundance of microbial flora in gastric cancer tumor tissues with high and low PLVAP expression. It was found that Fusobacteria and PLVAP expression showed a positive correlation in stomach cancer tissue [95]. In cholangiocarcinoma the expression of PLVAP is linked to DKK1/CKAP4/PI3K signaling, which are promoting angiogenesis [96]. Even in some non-neoplastic diseases, PLVAP is shown to be involved in the pathological progression of celiac disease (CD)-related liver injury. Circulating PLVAP is a marker of celiac disease-related liver damage and PLVAP expression was significantly higher in the blood of patients with celiac disease-related liver injury [97]. In an HCC mouse xenograft model, researchers observed tumor vascular thrombosis and large-area tumor necrosis following the injection of anti-PLVAP monoclonal antibodies, indicating a critical role for PLVAP [98]. These studies show that PLVAP has biological functions in normal and tumor cells.

1.5 The aim of this study

The aim of the present study is to explore the clinical value of PLVAP in HCC patients. Therefore, in this study, we attempt to investigate the role of PLVAP expression in HCC patient samples and its potential clinical implications.

Through the combined use of Lasso and SVM-RFE, critical markers, including PLVAP, have been identified for HCC tissue. Sets of genes are analyzed for the diagnostic and prognostic potential in HCC. Finally, the expression of PLVAP in cancer (tumor) and

para-cancer (normal) specimens from HCC patients is examined by immunohistochemistry and correlated to the patients' clinical information.

2. Materials and Methods

2.1 Materials

HCC tissue specimens

HCC tumor tissues and corresponding adjacent normal tissues were collected from Biobank from the Department of General, Visceral and Transplantation Surgery, LMU University Hospital, LMU Munich. All tissues were embedded in paraffin and stored at room temperature. This study was approved by the Institute of Ethics, History and Theory of Medicine, Faculty of Medicine, LMU Munich, Germany (approval reference number 24-0225). The inclusion criteria for patients are: 1. Patients are diagnosed with hepatocellular carcinoma; 2. Patients have no other types of tumors; 3. Patients need to have detailed clinical data; 4. Patients agree and are informed of the study.

Each tissue was sliced into 3 μm thickness for subsequent HE and IF staining. At least three areas of each slice were randomly selected for photography. For HE staining, one area was photographed at 100x, 200x, and 400x. For IF staining, one area was photographed at 200x and 400x.

Chemicals

Chemicals	Company	ID
Xylene	Carl Roth	9713.2
99% Ethanol	PanReac AppliChem	0v013438
Mayer's Hemalum Solution	Sigma-Aldrich	HX28488949
Hydrochloric acid(Hcl)	Merk	HC04270557
Eosin	Merk	E4382-25G
Neutral Balsam Mounting Medium	Sigma-Aldrich	HX04324861
Citric Acid Monohydrate	Carl Roth	Nr.3958.2
Trizma Base	Bio-Rad	#1610716
Sodium chloride(NaCl)	PanPeac Applichem	21010643
Tween 20	Sigma-Aldrich	PB79
Bovine Serum Albumin(BSA)	Carl Roth	Nr.0163.4
DAPI	Abcam	Nr.104139

Antibody

Antibody	Company	ID
PLVAP	Cell signaling	#38238
Goat anti-rabbit	Invitrogen	Alexa Fluor Plus 647

Instruments

Device	Company
Water bath	Memmert
Electronic balance	Waagen Dienst
Pipette	Eppendorf
Electronic pH meter	Knick Porlmamess
Shaker	Edmund Buehler
Refrigerator	Siemens
Microscope	Olympus
Fluorescence microscopy	ZEISS

Software

Software	**Company or institution**
R 4.2.1	R Foundation
Graphpad Prism 9	GraphPad Software
Image J 1.5.0	National Institutes of Health

Solution

Citrate Buffer	
Citric acid monohydrat	2.1 g
Distill water	1 L
pH	6

10x TBS	
Trizma base	24.2 g
Nacl	80 g
Distill water	1 L

1x TBS	
10x TBS	100 ml
Distill water	900 ml

1x TBS-T	
10x TBS	100 ml
Distill water	900 ml
Tween 20	1 ml

Blocking solution

1x TBS	1 ml
BSA	0.05 g

1% HCl-70% Ethanol	
70% Ethanol	118.75 ml
Hydrochloric acid 1 mol/L	68.75 ml

1% Eosin	
Eosin	2 g
96% Ethanol	200 ml

2.2 Methods – Protocols

2.2.1 Haematoxylin and Eosin Staining

After Paraffin sections were baked under the following conditions. 60° for 15 minutes. The sections were then immersed three times for 15 minutes in xylene. Then Specimens were removed from xylene and placed in 99% ethanol twice for 5 minutes each. 96% ethanol two times for 5 minutes each. 70% ethanol two times for 5 minutes each. Sections was washed in distill water for 1 minute. Then soak sections in Mayer's Hemalum Solution for 6 minutes. Wash sections in distill water for 3 seconds and then soak in 1% HCl-70% Ethanol for 3 seconds. Sections were removed and wash sections in running tap water one time for 5 minutes. Be careful not to wash sections directly, wash as gently as possible. The next step is sections were immersed in Eosin solution for 6 minutes, and then washed sections in running tap water for 30 seconds. Sections were placed in 70% ethanol for 1 minute, 96% ethanol for 5 minutes, and 99% ethanol for 5 minutes. Sections were placed in xylene two times for 5 minutes each later. After drying in air for 5 minutes, add 2 drops of Neutral Balsam Mounting Medium onto the tissue and then cover the slide with a coverslip, taking care not to leave any bubbles. At last, store the slices at room temperature.

2.2.2 Immunofluorescence staining

Day 1. First, preheat the water bath to 95°. Then The slices were immersed in xylene

twice, 15 minutes each time and in 99% ethanol, 99% ethanol, 96% ethanol, and 70% ethanol, 5 minutes each time. The next step is to wash the slices twice with 1x TBS, 2 minutes each time. The following step is that preparing an iron box and pour 1x citrate buffer (pH=6) into the antigen repair solution. The slices were placed in the antigen repair solution. The iron box was placed in a 95° water bath for 20 minutes. Iron box was removed from the water bath and cool at room temperature for 45 minutes. Then wash the slices twice with 1x TBS, 2 minutes each time. The slices were placed the slices in a humid chamber, drop 200ul of blocking solution on each slice, and incubate at room temperature for 45 minutes. At last, place the slices in a humid chamber, add 200ul of PLVAP antibody diluent (dilute PLVAP antibody with blocking solution at 1:200) to each slice, and incubate at 4° overnight.

Day 2. First, wash the slices twice with TBS, 2 minutes each time. Then place the slices in a humid chamber, add 200ul of goat anti-rabbit antibody diluent (dilute goat anti-rabbit antibody with blocking solution at 1:200) to each slice, and incubate at room temperature for 45 minutes. The following step is to wash slices twice with TBS, 2 minutes each time. Then, add 2 drops of DAPI to each slice, incubate at room temperature for 30 minutes, and then cover the slices with coverslips. (Note no bubbles) At last, store the slices at 4°C.

2.2.3 Bioinformatics data sources

2.2.3.4 Gene Expression Omnibus (GEO) database

GEO is a gene expression database established and maintained by the National Center for Biotechnology Information (NCBI) in the United States in 2000. So far, the GEO database is among the largest gene expression databases in the world, containing high-throughput sequencing data from all over the world. The data types include not only chip data, but also second-generation sequencing data and single-cell sequencing data. Species include not only humans, but also other mammals and plants. Disease types are not limited to tumors, but also many non-tumor diseases.

Because of its rich data and free global access, the GEO database is widely used and cited by many high-impact papers [99, 100].

HCC-related transcriptome sequencing data were retrieved from the GEO database. The GSE60502, GSE87630 and GSE76427 datasets include expression sequencing data of HCC tumors and normal liver tissues. Finally, these three datasets from GEO were selected for analysis (GSE60502 18 normal versus 18 HCC samples, GSE87630 30 normal versus 64 HCC samples, GSE76427 52 normal versus 115 HCC samples).

2.2.3.5 The Cancer Genome Atlas (TCGA) database

The Cancer Genome Atlas (TCGA) was established in 2006 through a collaborative effort between the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) and is a database focused on tumor research. So far, TCGA contains 33 tumor types and tumor-related data of more than 20,000 patients. The TCGA database provides RNA sequencing data, proteomics, methylation levels, non-coding RNA of tumor patients, and contains detailed clinical information of patients, such as age, tumor stage, etc. Compared with the data in the GEO database, the data in TCGA has been corrected by NCI and NHGRI, so the data quality is relatively high. TCGA has also been cited by many high-impact papers [101, 102].

2.2.3.6 Genotype-Tissue Expression (GTEx) database

GTEx is supported by the National Institutes of Health (NIH). By collecting tissue samples from normal donors, we try to understand the gene expression levels of different organs. The GTEx database contains DNA and RNA sequencing data of 54 different tissues and organs from 838 normal donors. Because the data of GTEx comes from normal people, GTEx only has gene expression data of normal tissues, which is also the biggest difference from other database such as GEO and TCGA databases. Apart from studying the differential genes between different tissues of the normal human body, another use of GTEx is to study tumors in conjunction with the TCGA database. By combining expression data from the GTEx database in order to

expand the sample size of normal tissues, the number of samples of tumors and normal tissues reaches a more credible level. Many high-impact factor literatures have adopted this method to study tumor diseases [103, 104].

In this study, in order to further increase the sample-years of normal liver tissue, we took a combined TCGA and GTEx approach to study HCC. UCSC XENA (https://xenabrowser.net/datapages/) provides RNAseq data in TPM format for unified processing of TCGA and GTEx. The gene expression data corresponding to HCC and the normal liver tissue corresponding to GTEx were extracted from UCSC XENA.

2.2.3.7 Single-cell sequencing analysis

We used the Tumor Immune Single Cell Hub (TISCH) single-cell database results to investigate protein expression in different cell types inside the tumor microenvironment. TISCH is a single-cell sequencing database focused on the tumor microenvironment. So far, the database contains 190 datasets on 50 tumors and 6,297,320 cells. Some high-impact factor papers also used this database for single-cell sequencing analysis [105-107]. Here, we examined the precise gene expression in the tumor microenvironment using the GSE166635 (HCC n = 2) dataset.

2.2.4 Exploration of HCC differentially expressed genes (DEGs) and screening of key genes

We examined the differences in gene expression between HCC tumors and healthy livers using R software. R software was first developed by statisticians Ross Ihaka and Robert Jetman in 1993 and is now mainly updated by the R Foundation. Because of its open source and free features, R software has been widely used in the field of bioinformatics. GSE60502, GSE87630, and GSE76427 gene expression matrices were created using R software (4.2.1).

The "limma" program was utilized to execute log2-transformation and background correction on the expression profiles. The Limma program is a widely used software package for assessing gene expression. It has strong data reading ability and

overcomes the problem of insufficient design for small samples. It is based on VOOM algorithm, which can perform differential analysis on chip data and transcriptome high-throughput sequencing data. Currently, it has been cited more than 16,000 times in the literature [108].

The transcriptome raw data we downloaded may come from different literature sources. These raw data have experimental errors caused by different times, operators, reagents, and instruments, which are reflected in batch effects on gene expression levels [109]. If we directly merge and analyze data from different studies, it will have some influence on the correctness of the analysis results. The SVA package is a software package that removes batch effects. By reducing differences between batches, we try to recombine data from multiple batches as much as possible. The SVA package supports to adjust batch processing and latent variables in prediction problems using the SVA function for proxy variable estimation, comBat function to directly adjust known batch processing effects, and the fsva function [110]. Batch effects correction across the two datasets, GSE60502 and GSE87630, was performed using the "SVA" software. These two datasets GSE60502 and GSE87630 were merged for further analysis. To construct a "heatmap" and a "volcano plot" of DEGs, the "pheatmap" and "ggplot2" packages were used.

2.2.5 Key genes screened from DEGs using machine learning methods

In HCC, machine learning techniques were used to filter important genes from DEGs. LASSO regression analysis was carried out using the "glmnet" package in order to find out important genes associated with HCC. Another machine learning approach to look for important genes from DEGs is SVM-RFE. We take the intersection of the important genes found by the two machine learning techniques using the Venn diagram.

2.2.6 Explore the diagnostic value of key genes and validate them in other datasets

Here, we assessed the diagnostic utility of important genes in HCC using diagnostic

Receiver Operating Characteristic (ROC) curves. The ROC curve is an important tool for evaluating binary variables, especially in medical diagnosis. The main process is to extract the expression of genes in normal liver and HCC tumors for each participant, and use the 'pROC' package to draw the ROC curve [111]. The Area Under the Curve (AUC) value can be obtained while creating the ROC curve using the 'pROC' package. AUC is used to evaluate the diagnostic value. The closer it is to 1, the better the efficacy. AUC > 0.80 means the diagnostic value is very good [112]. We used the combined expression matrix of GSE60502 and GSE87630 to evaluate the diagnostic value of key genes. At the same time, we used GSE76427, TCGA/GTEx about HCC to further validate the diagnostic value of key genes.

2.2.7 Evaluation of the prognostic value of key genes

In addition to diagnostic analysis of key genes, we carried out prognostic analyses of key HCC genes. To examine the prognosis of key genes in HCC, the Gene Expression Profiling Interactive Analysis (GEPIA) database was utilized. A large quantity of RNA sequencing data from TCGA and GTEx is integrated by GEPIA. So far, more than 5,000 papers have cited this database [113]. We assessed whether the groups with high and low gene expression differed in terms of Overall Survival (OS). The results were presented in the form of Kaplan-Meier (KM) survival curves.

2.2.8 Single-cell sequencing analysis in HCC tissues

The tumor microenvironment is made up of a variety of cell types, and the relationships between cells are complex. To understand the particular expression of genes in each type of cell, we used the single-cell sequencing approach for research. We employed the single-cell sequencing function of the online platform Tumor Immune Single Cell Hub (TISCH), and the selected data set was GSE166635. GSE166635 contains single-cell sequencing sets of 2 HCC patients.

2.2.9 Correlation analysis between genes and other markers (CD31, CD34, CD105 and VEGFR2)

Cluster of Differentiation 31 (CD31) is also called Platelet Endothelial Cell Adhesion Molecule-1 (PECAM1). CD31 is mainly expressed in vascular endothelium. It is crucial to the process of angiogenesis [114].

CD34 plays a role in the development of new blood vessels and is mostly expressed in endothelial progenitor cells. CD31 and CD34 are biomarkers of vascular endothelial cells [115].

The primary cells that express endoglin (CD105, ENG) are activated vascular endothelial cells and some hematopoietic cells. According to reports, increased CD105 expression in the tumor microenvironment enhances vascular endothelial cell proliferation and migration, as well as angiogenesis. In addition, the aberrantly networked blood arteries created by elevated CD105 expression have a high permeability, which facilitates tumor cell movement and metastasis [116].

Vascular Endothelial Growth Factor Receptor 2 (VEGFR2, KDR) is mostly expressed in vascular endothelial cells. VEGFR2 is the receptor for VEGF. When the ligand binds to the receptor, it activates the tyrosine kinase activity, which then initiates a number of signaling cascades, eventually stimulating the proliferation of vascular endothelial cells and playing a key role in angiogenesis and controlling vascular permeability [117]. The target of many anti-vascular drugs is VEGFR2. They inhibit angiogenesis by blocking the binding of VEGFR2 and VEGF, inhibiting tyrosine kinase activity, and blocking downstream pathways [118].

Here we try to analyze the expression correlation between genes and CD31, CD34, CD105 and VEGFR through the GEPIA database. The results are shown as a scatter plot, and the correlation between two variables is judged by the r value. $|r| \geq 0.7$ is highly correlated, $0.4 \leq |r| < 0.7$ is moderately correlated, $0.2 \leq |r| < 0.4$ is weakly correlated, and $|r| < 0.2$ is considered uncorrelated. Among them, $r > 0$ is a positive correlation, otherwise it is a negative correlation [119].

2.3 Statistical analysis

GraphPad prism 9 is a statistical software developed by GraphPad Software company. In this study, we used unpaired Student's t-tests to compare the two groups. Three or more groups were compared using one-way ANOVA with Dunnett's post-hoc test or Kruskal-Wallis analysis. Statistical result significant are indicated by * for $p < 0.05$, ** for $p < 0.01$.

3. Results

3.1 Screening out key genes in HCC using machine learning

I performed differential gene expression analysis of a merged HCC data set (GSE60502 and GSE87630) consisting of a total of 82 tumor samples and 48 normal samples. The filtering criteria for differentially expressed genes was the gene expression change fold is either less than 0.5 or larger than 2. Specifically, the change fold of gene expression is more than 2, which means that the gene expression in tumor tissue is at least 2 times higher than that in normal tissue. These genes are highly expressed genes in tumor tissue. The change fold of gene expression is less than 0.5, which means that the gene expression in normal tissue is at least 2 times higher than that in tumor tissue, that is, these genes are low-expressed genes in tumor tissue with an adjusted P value of less than 0.01. The results of differential expression analysis revealed an up-regulation of 148 and down-regulation of 391 genes in this Tumor versus Normal data set of HCC. In the heat map of this data set a clear difference in the expression of gene can be found. Further analysis on the DEG were performed to identify potentially characteristic and potentially clinically relevant genes (Figure 2A-E). For this we employed two different bioinformatic algorithms (1) SVM-RFE and (2) Lasso.

Machine learning SVM-RFE was used to screen for characteristic genes in the data set. For this the 'e1071', 'kernlab', and 'caret' packages were used. To ensure consistency in the results of each run, the random number seed was set to 123. The point corresponding to the minimum cross-validation root mean square error (RMSE) is taken as the optimal number of core genes. The result of this SVM-RFE analysis identified 28 genes (Figure 2C).

In the second technique, we employed a machine learning Lasso regression analysis to look for key genes in the same dataset. For this we used the 'glmnet' package. By using the 'glmnet' package ("binomial", alpha = 1, nfolds = 10) to perform lasso cross

validation, a lambda (λ) value was calculated. After the Lasso regression analysis, a set of 22 genes were identified as important genes (Figure 2D.

Finally, I compared the two independently calculated gene set and found an overlapping ("shared") set of 7 gene that were identified as characteristic by both methods (Figure 2E). Either analysis method also revealed unique genes, SVM-RFE (n = 21) and Lasso Algorithm (n= 15). We than used the "shared" gene set consisting of C-Type Lectin Domain Family 4 Member M (CLEC4M), Extracellular Matrix Protein 1 (ECM1), Dopamine Beta-Hydroxylase (DBH), Complement Factor Properdin (CFP), N-Acetyltransferase 2 (NAT2), C-X-C Motif Chemokine Ligand 14(CXCL14) and Plasmalemma Vesicle Associated Protein (PLVAP) for further analysis (Figure 2E).

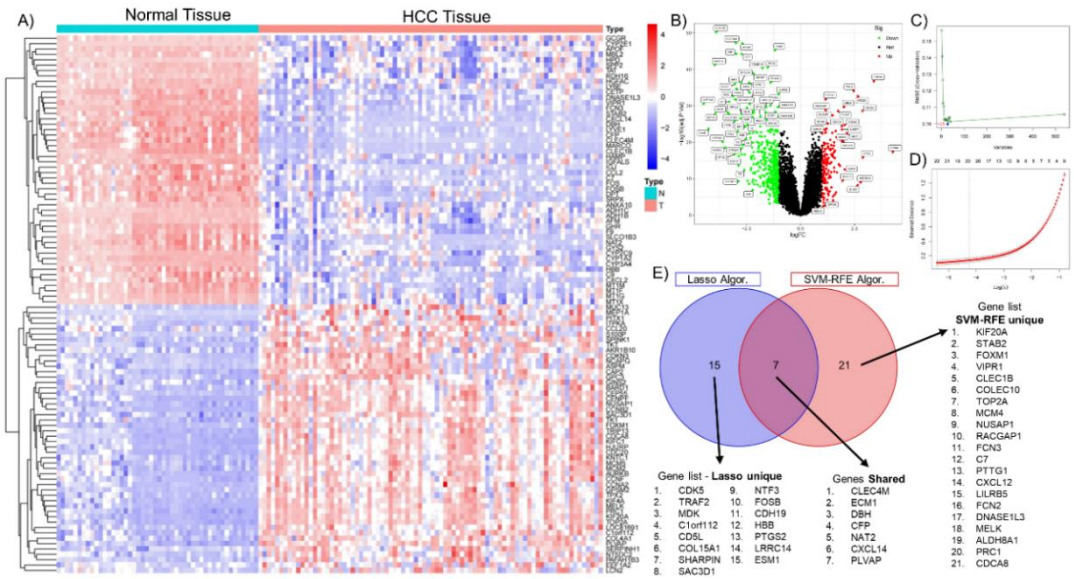

Figure 2. The process of screening key genes in HCC. A. After merging GSE60502 and GSE87630, differential expression analysis was performed between HCC tumors and normal livers. Heat maps are displayed as differential expression results, with the color depth indicating the strength of differential expression. The darker the color, the greater the difference. The color red indicates high expression, while the color blue indicates low expression. B. The differential expression results were displayed in a volcano plot. The color red represents high expression, while the color green represents low expression, while black indicates no significant difference. C. 28 genes were considered key genes in HCC by using the SVM-RFE algorithm to analyze differentially expressed genes. D. Genes with differential expression were analyzed using Lasso algorithm, and 22 genes were considered to be key genes in HCC. E. The 7 genes are considered as the shared part of SVM-RFE algorithm and Lasso algorithm,

and the results are displayed using a Venn diagram.

3.2 Diagnostic value of key genes in HCC

To better understand and evaluate the diagnostic relevance of the identified shared gene set we performed a ROC curve analysis of the gene expression matrix. Done by reading the gene expression matrix containing normal tissues and HCC tissues, and the calculation of ROC curves for each gene. The analysis looks ats the area under the curve (AUC) and confidence interval (> 95%) for key gene to estimate its relevant.

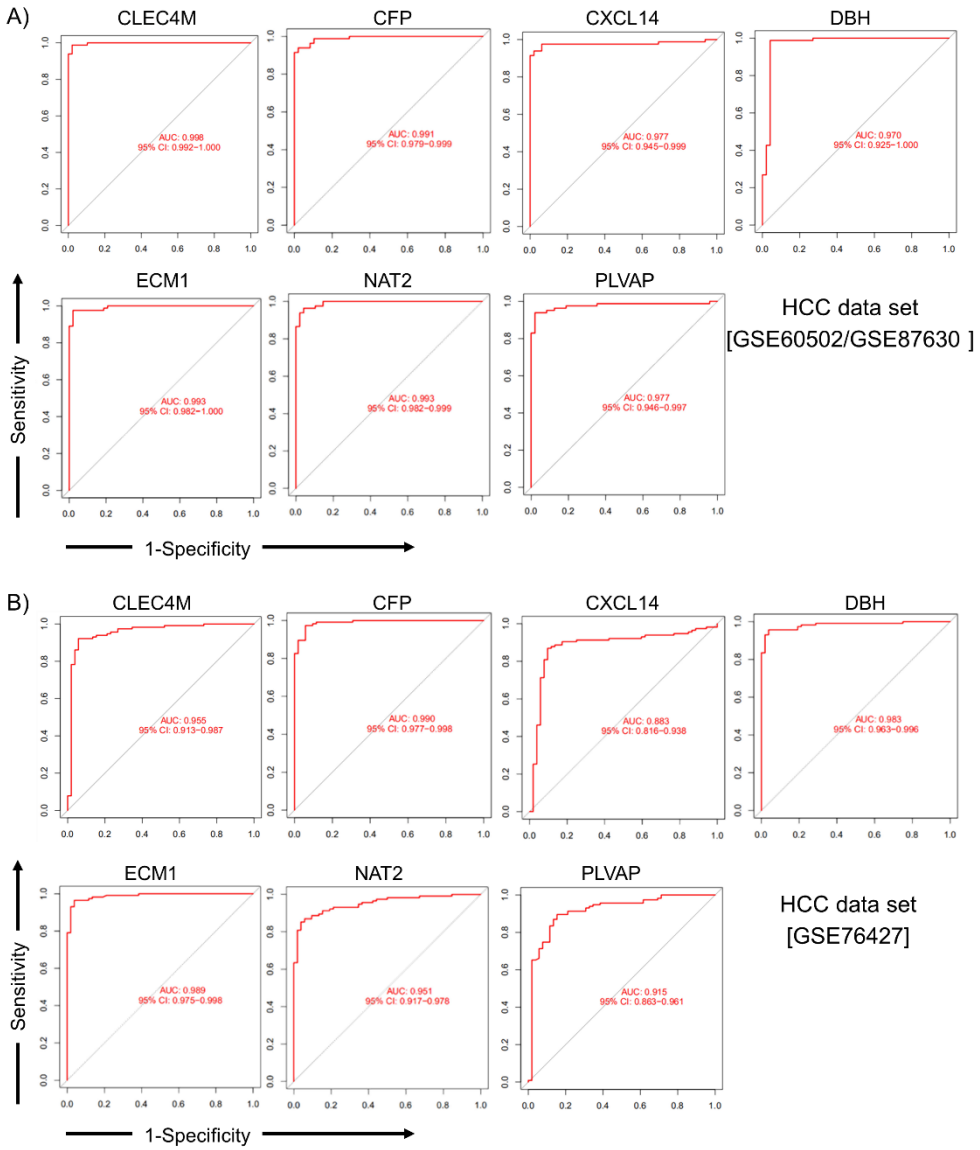

Figure 3_1. The diagnostic value of shared genes in HCC. A. Analysis of the diagnostic value of 7 shared genes in the combined liver cancer expression profile dataset GSE60502/GSE87630. B. Analysis of the diagnostic value of 7 shared genes in the

liver cancer expression profile dataset GSE76427.

The results from GSE60502/GSE87630 show an AUC value for the 7 genes of greater than 0.8 (Figure 3A_1 A), which indicates that these 7 genes could have a good diagnostic value. In order to see if these genes are only relevant in the GSE60502/GSE87630 HCC data set, we used another independent HCC sample dataset GSE76427(Figure 3A_1 B) and combined the gene expression matrix of TCGA and GTEx expanding the sample size for normal from TCGA/GTEx to 160 and for tumor from TCGA to 371 (Figure 3_2). We found that the 7 genes of interest scored ROC AUC > 0.8 in all HCC data sets. This data underlines the strength of these genes as potentially clinically relevant as they are identified as highly relevant for HCC.

The data from the combined GSE60502/GSE87630 gene sets showed that the 7 identified genes (CLEC4M, ECM1, DBH, CFP, NAT2, CXCL14 and PLVAP) correlated well with the diagnosis of HCC patients, which was corroborated by analyzing other sets such as the GSE76427 or the larger TCGA/GTEx sample set.

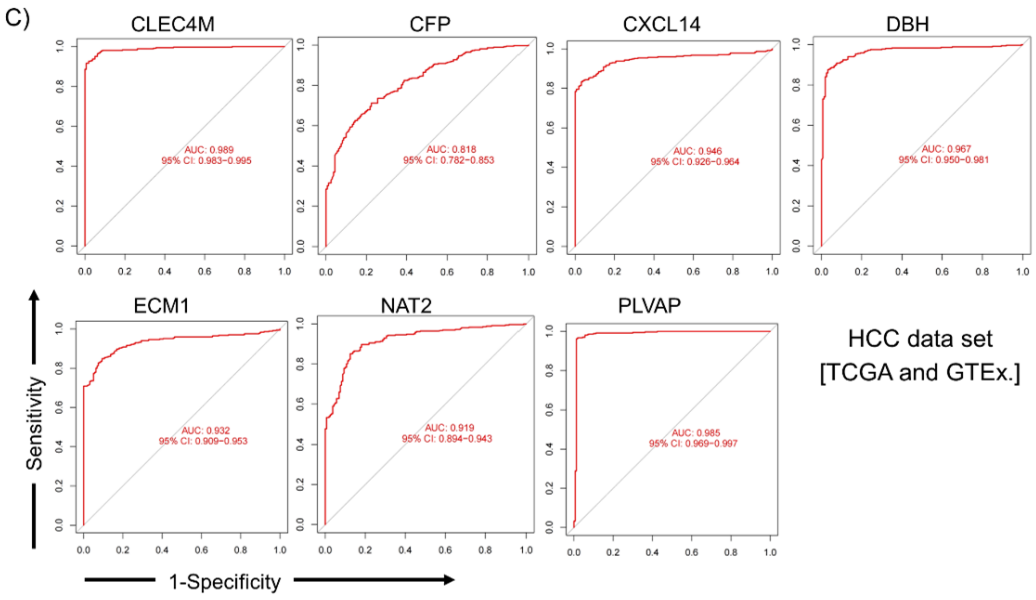

Figure 3_2. The diagnostic value of shared genes in HCC. C. Seven genes shared in the combined TCGA/GTEx liver cancer expression profile dataset are evaluated for diagnostic value.

3.3 Evaluation of the prognostic value of key genes

In order to further evaluate the potential use of the 7 genes of interest in HCC, for example as prognostic markers, we used the GEPIA (https://gepia2.cancer-pku.cn/) analysis tool. Here median gene expression values for each HCC patients were used to divide the patient cohort into two groups. Patients were categorized into two groups based on their gene expression levels: those with higher gene expression than the median value (high expression group; red) and those with lower gene expression than the median value (low expression group; blue). The prognostic outcomes of the genes between the patient groups with high and low expression were displayed as Kaplan-Meier (KM) survival curves (Figure 4). The *p*-value is used to determine whether there is a survival difference between patient groups, with a *p*-value < 0.05 indicating a survival benefit. The analysis revealed that most genes (CLEC4AM, ECM1, CXCL14, NAT2, DBH) did not strongly correlated with the survival of HCC patients. However, for PLVAP a strong correlation (LogRank = 0.027, and p-value < 0.05) with improved survival of HCC patients was found (Figure 4). This suggests that patients with HCC who express more PLVAP will be better than those who express less.

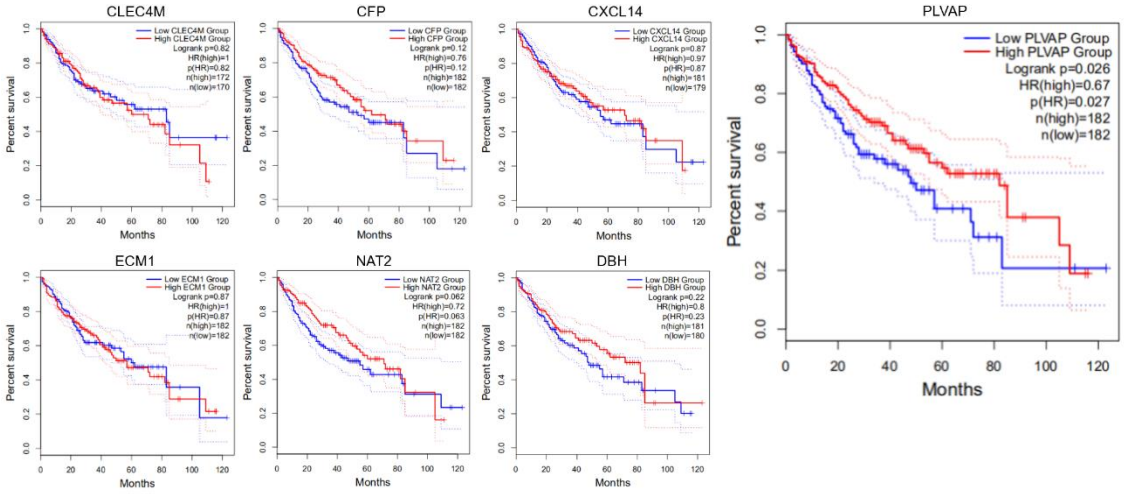

Figure 4. Prognostic analysis of 7 shared genes from GEPIA. Only PLVAP showed statistical significance in HCC (p<0,05) and HCC patients expressing high levels of PLVAP had better outcomes than those with low levels. Patient groups were divided into high and low expression groups based on the median value of gene expression. The colors red and blue indicate high expression and low expression, respectively.

3.4 Single-cell sequencing analysis of PLVAP in HCC tissue

Tumor tissues are made up of a complex framework of different cell types, such as stromal and immune cells. We investigated the PLVAP expression profile of various cell types using the GSE166635 HCC sample set (n=2). With the TISCH-tool 25,189 single-cell transcriptome data set was analyzed. We examined PLVAP expression in immune system, stromal, and tumor cells. The findings demonstrated that stromal cells expressed PLVAP at a significantly higher level than other cell types, such as immune or tumor cells (Figure 5A). A more detailed analysis of the different cell types was done, revealing that within the stromal cell population PLVAP expression was the strongest in endothelial cells (Figure 5B).

According to these results the observed PLVAP expression is originating from endothelial cell population, which is cell type found in the vascular system. Therefore, we investigated the PLVAP expression pattern of endothelial cells in HCC patient tissue as a potential prognostic tool/biomarker.

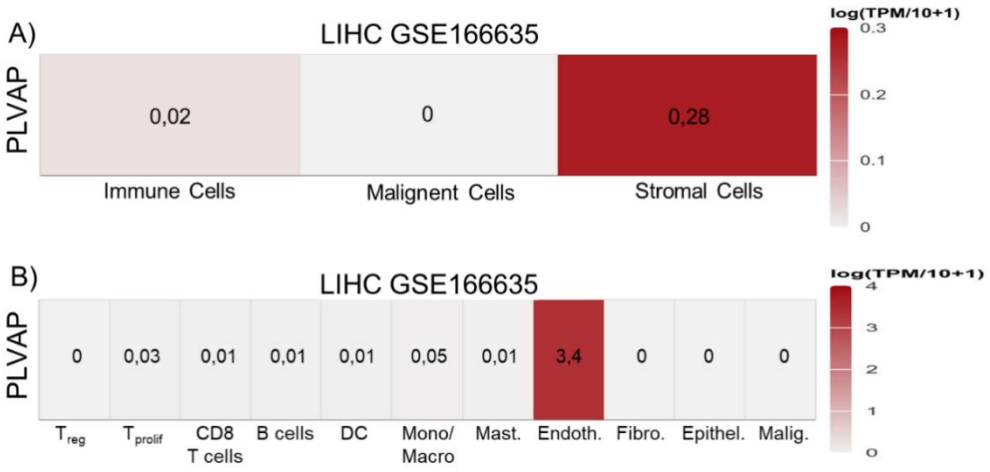

Figure 5. Single-cell sequencing analysis of PLVAP expression comes from GSE166635. A. Expression of PLVAP in tumor cells, stromal cells, and immune cells. B. The expression of PLVAP in various types of tumor cells, stromal cells, and immune cells. Depending on the depth of the red, the greater the expression. Immune cells: T cells (T regulatory cells, T proliferating cells), DC cells (Dendritic cells), Monocytes/Macrophages (Mono/Macro), Mast Cells (Mast)), Tumor cells; Stromal cells (Endothelial cells (Endoth), Fibroblasts (Fibro), Epithelial cells (Epithel)). Shown is gene expression as TPM (transcripts per million).

3.5 Correlation analysis between PLVAP and endothelial biomarkers (CD31, CD34, CD105 and VEGFR) in HCC

Several biomarkers, including CD31 (PECAM1), CD34, CD105 (ENG) and VEGFR2 (KDR), are expressed by endothelial cells [120]. We used the GEIPA tool to correlate the expression of the 4 endothelial markers with the expression of PLVAP in HCC samples (Figure 6). The results are shown in the form of correlation scatter plots for each of the endothelial markers. The analysis revealed a correlation coefficient of PLVAP and CD31 of R = 0.57, CD34 of R = 0.69, CD105 of R = 0.57, and VEGFR2 of R = 0.59 all with a *p*-value < 0.05 (Figure 6). A R coefficient of $0.4 \leq |R| < 0.7$ indicates a moderate correlated between two variables. Which means that PLVAP expression is correlated with the expression of endothelial biomarkers. Endothelial cells are specialized cell type and form the vascular system in the human body. The liver, and subsequently also HCC tumors are highly vascularized tissues [121]. These results point toward the vascular system as an origin for the observed PLVAP expression. Which could be used as a prognostic marker to stratify HCC patients.

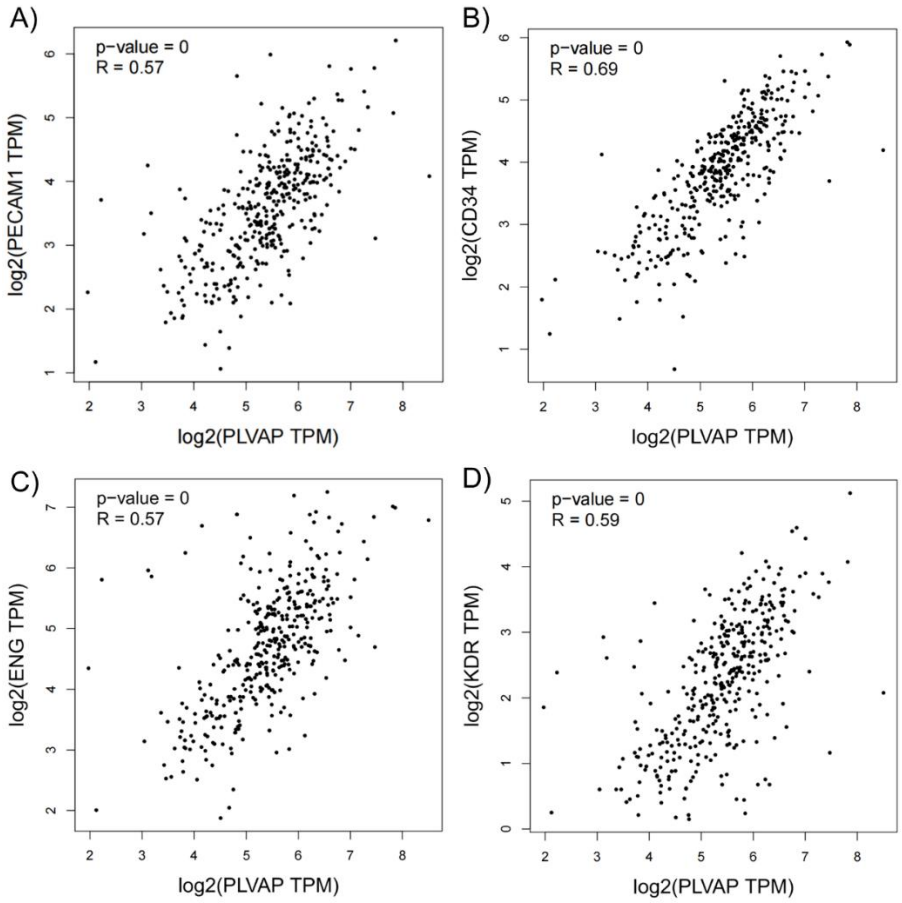

Figure 6. Correlation analysis between the expression of PLVAP and common endothelial cell biomarkers in HCC. A) Correlation analysis between PLVAP and CD31 (PECAM1), correlation coefficient R = 0.57. B) Correlation analysis between PLVAP and CD34, correlation coefficient R = 0.69. C) Correlation analysis between PLVAP and CD105 (ENG), correlation coefficient R = 0.57. D) Correlation analysis between PLVAP and VEGFR2 (KDR), correlation coefficient R = 0.59.

3.6 Hematoxylin-Eosin and immunofluorescence staining of HCC tissues

For the purpose of examining PLVAP expression in HCC patients' endothelial cells, we collected 19 paraffin-embedded sections from tumor tissue as well as corresponding non-tumor (normal) tissue. For each patient (tumor and normal) sections where HE stained to identify blood vessels and then PLVAP staining was determined using immunofluorescent histochemistry (Figure 7). For the analysis we aimed to image 5 randomly selected vascularized areas for each patient sample. However, for some samples we could only identify 3 vascular areas which were than used. The staining

for PLVAP revealed a strikingly clear difference between tumor a normal tissue (Figure 7).

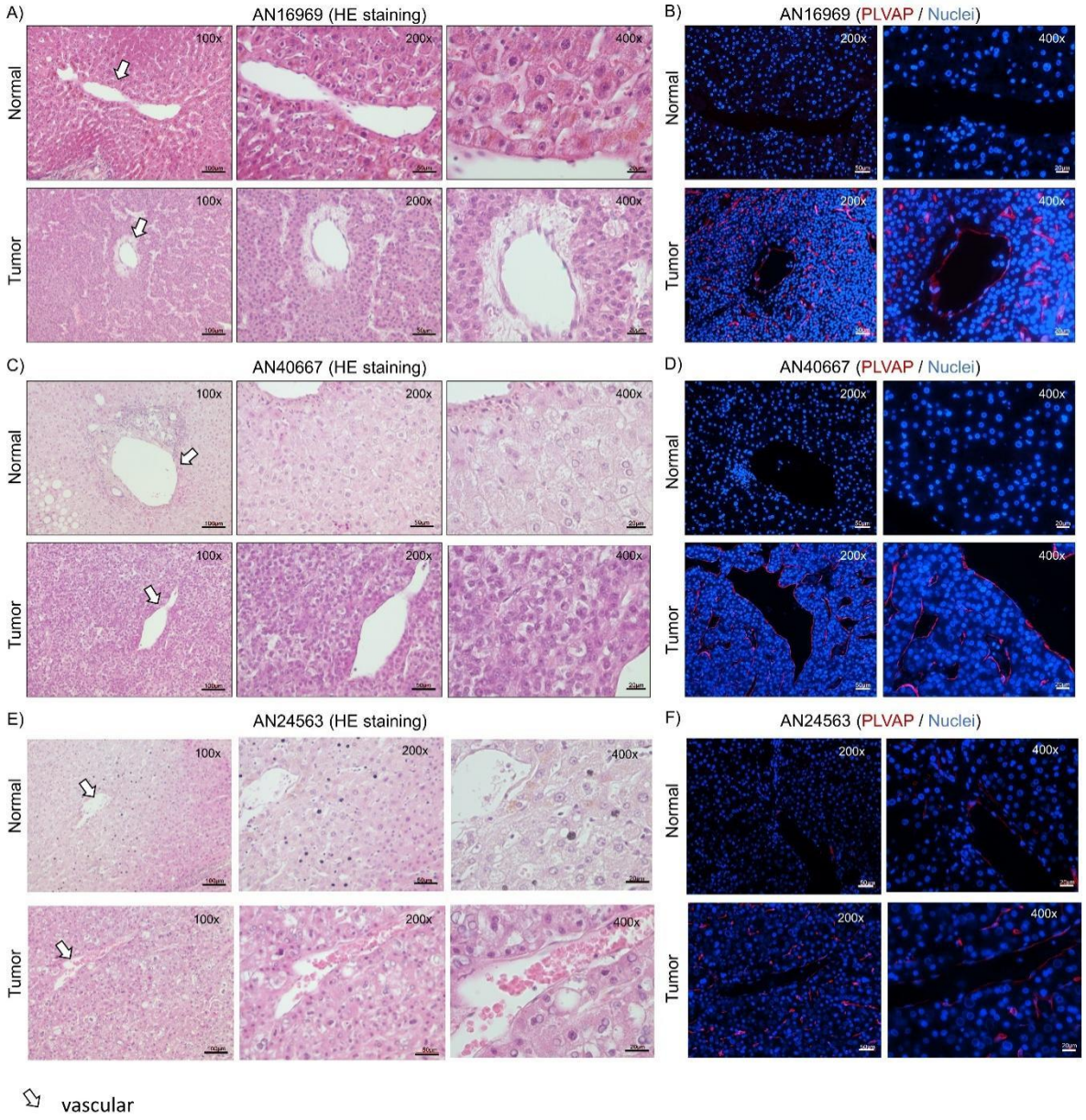

Figure 7. Tumor and normal tissue stained by HE and for PLVAP. Shown are 3 representative HCC patient samples. A, C, E) HE staining are shown, blood vessels are indicated (arrow). B, D, F) PLVAP expression is shown in red, with cell nuclei counterstaining (DAPI, blue). Scale bar and magnification are shown.

Even though blood vessel were clearly present in normal tissue sections, as identified by HE staining, the PLVAP signal was mostly very weak. In clear contrast, in all tumor

samples, we found a moderate to strong PLVAP signal in endothelial cells. Although no endothelial specific stating was done, the localization of the PLVAP signal i.e., at a blood vessel, as well as its longitudinal shape matches with the staining pattern expected/known form biomarkers in endothelial cells (Figure 7). Interestingly, we found in 3 cases a strong PLVAP expression despite no obvious vascular system. This could be due the presence of microvascular systems that are known to be more difficult to identify without additional markers. We separated these cases into a separate group (microvascular group). Further studies would be needed though to understand that PLVAP expression could be used to identify microvascular systems as well as its use a prognostic marker.

We can find the pathological differences between tumor tissue and normal liver tissue under the microscope. From the HE staining results, we can see that the cells of the adjacent tissue are closely arranged. The cell nucleus is usually centered and stained evenly. In the vascular structure area, flat endothelial cells can be seen under a high-power microscope. However, the liver cancer cells are arranged disorderly and are generally arranged in a disordered manner. The cell nucleus morphology is irregular, and the heterogeneity between the cell nuclei is obvious. The above characteristics are consistent with the pathological characteristics of tumors and normal tissues.

In HCC samples increased presences of stromal and endothelial cells are known, which could indicate angiogenesis via increased number of macro and microvascular system (Figure 5 and Figure 6). However, the precise function of PLVAP expression in cells of the microvascular system remains unclear, but due to the fact that microvascular systems are considered very small blood vessel a similar role could be assumed.

Following up on the observation of PLVAP expression in HCC endothelial cells we correlated the available clinical information for each patient with the PLVAP expression pattern.

3.7 Correlation analysis between PLVAP expression and clinical data

Based on the PLVAP expression we found in the 19 pairs of HCC and normal tissue samples, we grouped the patients into three groups 1) Moderate Expression Group, 2) High Expression Group and 3) Microvascular Group. The cut-off for the Moderate and high expression group was a PLVAP staining value of 200. These groups reflect that there was no tumor sample that did not show significant PLVAP expression (Figure 8A).

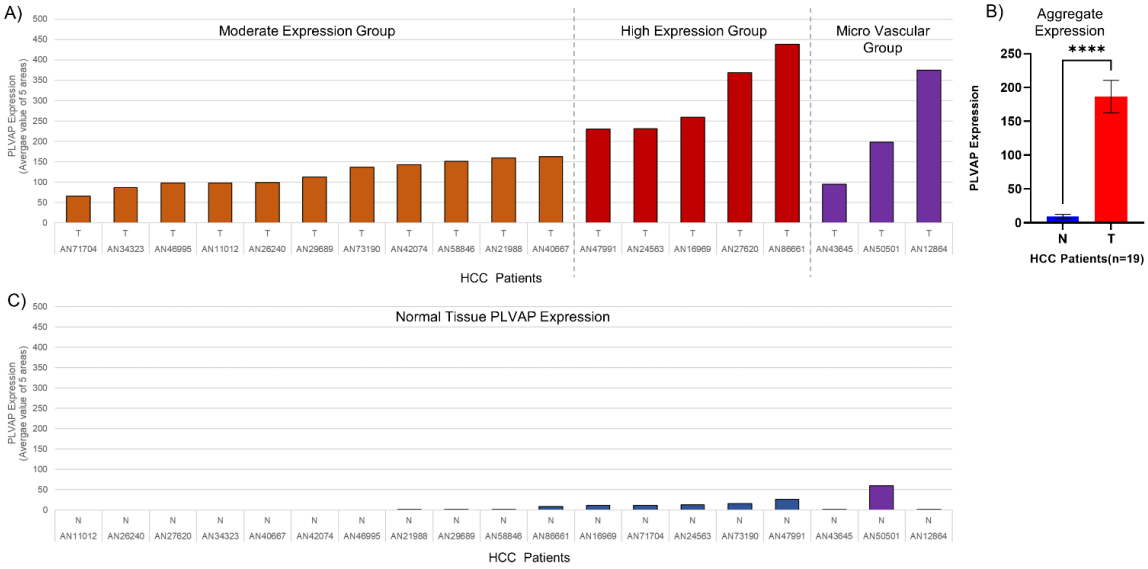

Figure 8. Summary of PLVAP expression in HCC patient samples. A) Expression level of PLVAP in HCC tumor tissues. Staining intensity of 3-5 random vascular areas are shown. Groups are based on expression cut-off < 200 (moderate and high) and for samples with PLVAP staining but no visible vascular system (microvascular group, purple). B) Aggregate PLVAP expression shown for 19 HCC patient sample. SEM indicated, **** indicates significances T-test *p*-value <0.001. C) PLVAP expression of normal tissue from corresponding HCC patients.

The three patient samples that showed PLVAP expression but had no obvious vascular structure were also separate, as these do not match the criteria for the current study aim i.e., PLVAP expression in vascular areas. The significant difference in the PLVAP expression is also illustrated in the aggregated PLVAP expression where all 19 samples are combined (Figure 8B).

Although we observed a generally strong PLVAP expression in HCC tumor samples, some are still stronger, which lead us to our sample grouping (moderate vs high).

Following the PLVAP expression analysis (Figure 8) we correlated the both groups with respect to the available clinical patient information's (Table 1).

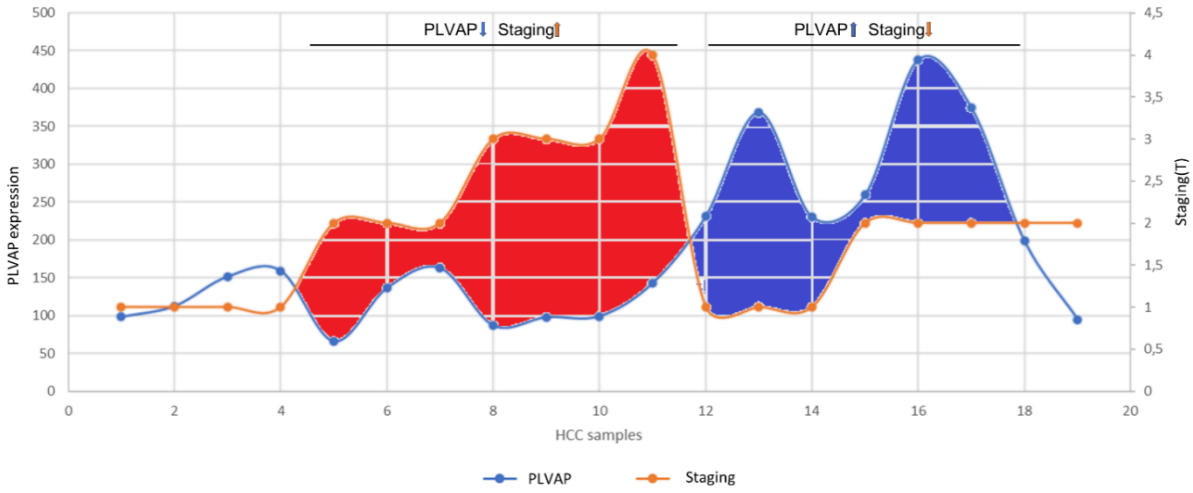

Figure 9. Correlation analysis between PLVAP expression and T stage in HCC patients. Each HCC patient's PLVAP expression level is represented in blue, and the T stage is represented in orange.

The difference in M/F ratio is most likely due to the low number of high PLVAP expressing samples, though it could indicate a gender specific difference. Another difference between both groups is the HCC patients T staging. Interestingly, patients with higher T stages (up to IV), which correlates with a more aggressive tumor, are in the moderately expressing PLVAP group. For the most part high PLVAP expressing samples had lower T stages (I or II) (Figure 9). However, due to the low number of patients, particular in the high PLVAP expression group final conclusions are difficult to draw. But it is clear that PLVAP expression found in HCC samples is driven by endothelial cells, and that there is a potential correlation between T stage and strength of PLVAP expression.

Table 1. Clinical information on PLVAP expression groups (moderate, high and µVascular)

	PLVAP Expression		
	Moderate	High	µVascular
Sample (n)	11	5	3
Diagnosis	HCC	HCC	HCC
Median Age (years)	63	63	71
Error (+/- years)	9,6	9,8	4,6
PLVAP	112,3	259,3	198,2
Ratio (M/F)	1,75	4	3
Chemo. Treat.	no	no	no
Cirrhosis (Y/N/na)	3/7/1	0/4/1	1/2/0
Fibrosis (Y/N/na)	3/5/2	2/3/0	3/0/0
Steatosis (avrg.)	6%	8%	17%
Pathology (T)	T1 (4), T2 (3), T3 (3), T4 (1)	T1 (3), T2 (2), - -	- T2 (3), - -
Pathology (L)	L0	L0	L0
Pathology (G)	G1 (1), G2 (7), G3 (3)	- G2 (4), G3 (1)	- 3(G2), -
Pathology (UICC)	I (2), Ia (1), Ib (1), II (2), IIIa (3), IIIc (1), IV (1)	I (2), Ia (1), II (2), - -	- II (2), IIa (1) - -

4 Discussion

4.1 Background of HCC

Liver cancer is one of the most frequent human malignant tumors, with HCC being the most common subtype [122]. With the third-highest death rate and the sixth-highest tumor incidence worldwide, HCC represents a significant issue for global health [123]. Results from the latest epidemiological surveys show that the incidence and mortality of HCC are highest in Asia, especially in East Asia [124]. Risk factors for HCC include HBV, HCV, Nonalcoholic fatty liver disease (NAFLD), alcoholism, and aflatoxin. The risk factors for HCC are evolving as a result of decades of extensive promotion of the hepatitis vaccine as well as changes in the diet of people and lifestyle choices. The incidences of HCC related to HBV and HCV will decrease, while HCC caused by NAFLD will increase rapidly and may exceed incidences of viral hepatitis in the next decade [125]. Studies predict that by 2030, the annual incidence of HCC caused by NAFLD will increase by 45% - 130%. It is expected that NAFLD will become the primary risk factor for the development of liver disease [126, 127]. NAFLD is gaining more and more attention from society as well as the development of associated diseases. Improving the understanding of the underlying modifications and the pathogenesis are growing research topics.

As science and technology advance at a rapid pace, new techniques and markers for diagnosing HCC are being developed. The diagnostic methods for HCC generally include imaging techniques, analysis of serum markers, pathological examination, and molecular diagnostics. Due to the complexity of HCC as a disease, relying on only one existing diagnostic method is not advised, considering the limitation in sensitivity and specificity of each analysis. Therefore, there is a constant search for new and potentially better targets and more specific diagnostic methods.

With the continuous improvement in tumor research, great progress has been made in the treatment of HCC. Currently, available HCC treatment strategies include liver

resection, liver transplantation, interventional therapy options like radiofrequency ablation or transarterial chemoembolization, targeted therapy, and immunotherapy. Although there is an array of treatment options available, each method has specific indications and limitations. How to reasonably combine two or more different approaches to make up for the shortcomings of each method and develop new and more precise treatments is an object of current and long-term clinical research efforts.

4.2 Lasso and SVM-RFE Machine Learning

The rapid growth of bioinformatics has opened up new avenues for a more in-depth understanding of illness causation. Bioinformatic analysis can quickly identify targets through large-scale analysis of high-throughput sequencing data, and help to focus research on the most promising targets. In recent years, popular machine learning and artificial intelligence algorithms have also been widely used in bioinformatics. By employing such analysis methods large scales of disease-associated genes and molecules were accurately and quickly screened. The constantly growing and wide range of bioinformatics data sets also increases the accuracy and reproducibility of the results, which is an advantage that is expected to get even better in the future. Globally, some countries have initiated and established data centers, where data sets are provided for public research efforts. Researchers can use such data sets for their respective analysis. Currently, bioinformatic plays an important role in screening for potential targets. However, the need for experimental confirmation of identified targets remains. However, due to the screening of many targets, the research effort can be focused on the most likely candidates.

In order to identify new potentially relevant genes of HCC, we applied a Lasso and SVM-RFE machine learning approaches on a large HCC data set. From this effort, we were able to identify a list of 7 genes (CLEC4M, ECM1, DBH, CFP, NAT2, CXCL14, PLVAP).

The main functions of CLEC4M (C-Type Lectin Domain Family 4 Member M) are cell adhesion and pathogen recognition. Recognized pathogens include an array of viruses

such as Ebola virus, hepatitis C virus, and influenza A virus [128]. CLEC4M overexpression is shown to inhibit liver cancer cell growth and promotion of apoptosis, and it has been linked to the Janus kinase 1/signal transducer signaling pathway [129]. The expression of CLEC4M diminishes with the advancement of HCC [130].

For multiple tumors, ECM1 (Extracellular Matrix Protein 1) expression has been shown to be involved in tumor proliferation and invasion [131-133]. In HCC, ECM1 appears abundantly expressed and is linked to vascular invasion and TNM staging. ECM1 also promotes HCC migration and invasion through epithelial-mesenchymal-transition (EMT) [134].

Growth of some tumors can be regulated by the nervous system. For DBH (Dopamine Beta-Hydroxylase) an involvement in the regulation of critical pathways was described, associating the central nervous system (CNS) and biosynthesis/production of norepinephrine as a key regulatory target. Activating the promoter of DBH can slow tumor progression [135].

In breast cancer, CFP upregulates DDIT3 expression and ultimately inhibits cancer cell growth [136]. A variation in the NAT2 gene is linked to the prevalence of smoking-related HCC [137]. Through inhibition of the Akt/mTOR signaling pathway, CXCL14 prevents HCC tumor cells from proliferating [138].

It was shown that HCC tissues were enriched in endothelial cells expressing PLVAP and tumor-associated macrophages expressing FOLR2. These cells transmit information through the VEGF-NOTCH signaling pathway, linked to the immune escape of HCC tumor cells [139].

With a single-cell sequencing and spatial transcriptomics analysis, it was discovered that PLVAP-expressing endothelial cells contributed to the immunosuppressive microenvironment in HCC [140]. In an animal model of HCC, anti-tumor effects were observed by injecting PLVAP antibodies, with minimal systemic adverse reactions, indicating that PLVAP could be a target with anti-tumoral activity [98].

Thus, of the described seven identified genes, only the genes CLEC4M, ECM1, NAT2, CXCL14, and PLVAP are reported to have some role in HCC either shown in in vivo or

in vitro conditions. For DBH and CFP no information linking their function to HCC was found. This indicates that the bioinformatic approach identified several genes that are already known to have a role in HCC as well as some that might be new targets.

4.3 Analysis of the diagnostic value of genes

We further explored the diagnostic value of these seven genes in HCC using three different datasets. The results indicated a potential diagnostic value, considering the differential expression of the genes in normal and tumor tissue.

When genes are considered as diagnostic markers for diseases, factors such as sensitivity, repeatability, clinical relevance, and cost-effectiveness need to be assessed [141]. In this study, we focused on PLVAP and its in vitro detection in tumor and normal liver tissue. Our results showed that PLVAP was mainly present in tumor vascular endothelial cells, and the expression level of PLVAP was negatively correlated with the T stage of HCC patients. However, this can be only considered an early step in the validation process for this promising target.

We also analyzed the available literature on the identified genes for connections to HCC. Only for ECM1, we were able to identify reports that discuss it as a potential diagnostic biomarker for HCC [142]. For none of the other genes, we were able to find similar reports. As we used RNA sequencing data sets for our machine learning approach , it is important to also investigate protein levels as well as localization, as there are several layers of regulation between RNA and functional protein [143]. Therefore, we opted for the immunohistochemistry method to detect PLVAP in tumor and normal liver tissue.

In summary, our results indicate that seven genes showed good diagnostic potential from RNA level, but whether the seven genes can be used as diagnostic markers for HCC in clinical requires more independent clinical cohorts and rigorous analysis and evaluation before a final conclusion can be drawn. We started with one of the genes in this study.

4.4 Analysis of the prognostic value of genes

We also examined the prognostic significance, i.e. the impact of the genes on the overall survival (OS) of HCC patients. We found that particularly PLVAP has good prognostic value. Furthermore, the difference in the expression of PLVAP in tumor (high) versus normal (almost none) tissue was quite interesting. In addition, the expression levels in tumor tissue, appears to correlate with the HCC tumor staging. We were able to separate the samples into several groups, according to the expression level. Samples from patients with better prognosis (Stages I and II) showed higher PLVAP expression compared to patients with worst prognosis (Stages III and IV). This is indeed an interesting observation and could help to identify patient groups with better prognosis by staining for PLVAP.

Previous studies have shown that CD105 and CD34 are negatively correlated with the TNM stage of HCC patients. CD105 and CD34 are more abundantly expressed in early HCC (Stages I and II) than in advanced HCC (Stages III and IV) [144, 145]. In this experiment, PLVAP was specifically expressed in tumor vascular endothelial cells, and we also observed, PLVAP to be negatively correlated with the stage of HCC patients like CD105 and CD34. The connection of PLVAP with endothelial cells and the formation of blood vessels might indicate that in HCC a high demand for vascularization is present earlier in the disease stages whereas later less so. In the early stages of tumor, the tumor and new blood vessels grow synchronously, but in the late stages of tumor, the tumor and new blood vessels grow asynchronously. The growth rate of tumor cells is much faster than the rate of angiogenesis, so the late-stage tumor tissue is in a state of hypoxia [146]. This might result in the observed PLVAP expression pattern.

4.5 Single-cell sequencing data

Tumor tissues contain a variety of cell types, and the relationships between them are highly complex. We explore the expression of PLVAP in the tumor microenvironment

using single-cell RNA sequencing tool from GSE166635 HCC sample. PLVAP is expressed in mainly stromal cells, and here predominately in endothelial cells. It appears not expressed in malignant cells, and at best very low in other cell types. The results suggested that PLVAP may be specifically expressed in endothelial cells.

4.6 Detection of PLVAP expression in HCC patients

For the expression analysis of PLVAP in HCC patient samples, we collected respective samples and did HE stain and IF. The staining of 19 pairs of normal and tumor tissue of HCC tissues showed that PLVAP was mainly expressed in the endothelial cells of blood vessels, which was consistent with the results from single-cell sequencing data analysis. At the same time, we found that in normal liver tissue, PLVAP was not expressed or at very low levels, even though for both tissue areas, vascularization was seen. In general, PLVAP is highly expressed in tumor vascular endothelium and very low in normal vascular endothelium.

Studies have shown that there are significant differences in gene expression, function, and morphology between tumor vascular endothelium and normal tissue vascular endothelium [147]. These differences are closely related to tumor angiogenesis, metastasis, and invasion. Compared with normal tissue, tumor tissue is often rich in vascular structures. These often also abnormally increased blood vessels provide essential nutrients supporting the need of tumor cells and their rapid growth. Genes involved in blood vessel development that are aberrantly expressed are frequently found in tumor vascular endothelial cells [148, 149]. With regard to PLVAP, researchers reported a strong expression in cholangiocarcinoma tumor tissue. They found that cholangiocarcinoma tumor cells secrete DKK1 (Dickkopf WNT Signaling Pathway Inhibitor 1) protein which in turn binds to CKAP4 (Cytoskeleton Associated Protein 4) on vascular endothelial cells surrounding the tumor cells. Downstream signaling pathways that are activated following this signaling cascade is the PI3K/Akt signaling pathway which in turn results in an increase in PLVAP expression. This is associated with the formation of new blood vessels [96]. Our data on the expression of PLVAP in

HCC tissue appears to follow similar expression patterns. Interestingly we found PLVAP expression elevated in tissue areas that do not show obvious vascular structures, which might indicate the presence of so-called microvascular systems.

Such micro vessels in liver tissue are also formed by hepatic sinusoids [150, 151]. Their main function is to allow the abundant hepatic sinusoids in the liver to contact liver cells, ensuring that there is enough space for blood and liver cells to exchange substances [152]. Although the lumen structure of hepatic sinusoids is difficult to observe under a microscope, they still have endothelial cells [153]. These endothelial cells are only loosely arranged, which allows the passage of large molecules, but also tumor cells. In contrast to properly formed tight blood vessels, this type of vessel can be associated with the metastasis of tumor cells [154]. In addition to playing a role in material exchange, endothelial cells also play an important role in pathophysiology and PLVAP expression could be used as a marker [155, 156].

Based on this result, we then performed a correlation analysis between PLVAP expression and vascular endothelial biomarkers (CD31, CD34, CD105, and VEGFR2) in HCC. CD31, CD34, CD105, and VEGFR2 all play an important role in the pathological progression of HCC, especially in the process of angiogenesis and tumor invasion of blood vessels [145, 157-163]. We explored the relationship between PLVAP and these common vascular endothelial biomarkers and found a strong correlation.

4.7 Correlation analysis between PLVAP and vascular endothelial biomarkers

We analyzed the correlation between PLVAP and CD31, CD34, CD105, and VEGFR2 using the correlation analysis function of the bioinformatics tool GEPIA in HCC. The results found that PLAVP and these endothelial cell biomarkers moderate positive correlation. CD31, CD34, CD105, and VEGFR2 are very important molecules in angiogenesis and vascular permeability. CD31 is mainly present on the surface of endothelial cells and its function is to maintain the integrity of the vascular endothelial structure [164]. CD31 is used as a biomarker of endothelial cells in angiogenesis experiments [165]. CD34 is mainly found on the surface of endothelial progenitor cells

[166], which would indicate a role in maintaining new blood vessels [166]. CD105 is mainly found on the surface of endothelial cells and is highly expressed in endothelial cells [167]. It is also a co-receptor of transforming growth factor β (TGF-β) [168]. It is also used as a biomarker of vascular endothelial cells in the study of angiogenesis experiments [169]. VEGFR2 is mainly found on the surface of endothelial cells and is a receptor for VEGF [170]. After VEGF binds to VEGFR2, a series of signaling cascades is activated which results in the proliferation of endothelial cells and angiogenesis [171]. VEGFR2 is also an important target for anti-tumor drugs like VEGFR2 inhibitors, such as Anlotinib, or Sorafenib [172]. Targeting angiogenesis is a highly effective therapy strategy.

PLVAP may also play a role in angiogenesis and vascular permeability processes. There is a clear correlation between PLVAP and common endothelial cell markers. Even though the single-cell sequencing data and immune tissue analysis indicate a general endothelial expression of PLVAP, we found almost no expression in normal tissue, which would limit the use of PLVAP as a vascular endothelial marker. This difference in the expression of PLVAP could be due to signaling within the tumor microenvironment. In normal tissue, these signaling ques appear to be less pronounced, and the endothelial cells remain not activated [96]. So far, there are no reports on a specific mechanism of high PLVAP expression in HCC, which, however, warrants additional studies to further explore this observation.

4.8 Clinical value of PLVAP in HCC

The clinical information of our HCC patient cohort showed that all HCC patients are middle-aged (50.25 ±4.573) to elderly (68.13 ±5.617). Studies have shown that the middle-aged and elderly are more likely to suffer from HCC as age is a known as a risk factor [173]. The damage to the liver caused by hepatitis virus, alcoholic liver disease, or non-alcoholic liver disease simply accumulates over time, and eventually results in the development of HCC [174, 175]. Among our 19 HCC patients we had more male patients than female ones. This is consistent with common epidemiological data

provided by SEERS (Surveillance, Epidemiology, and End Results) which reports a gender ration of 3:1 for (male:female) [176]. The Surveillance, Epidemiology, and End Results (SEER) Program provides information on cancer statistics in an effort to reduce the cancer burden among the U.S. population. This is most likely linked to unfavorable life style habits (smoking or drinking) of men compared to women [177]. There are also studies showing that estrogen has a protective function on the liver [178]. In this study, some of the 19 HCC patients had also liver cirrhosis and liver fibrosis. Cirrhosis, liver fibrosis, and HCC are closely related diseases. They develop due to long-term stimulation with risk factors, resulting in the necrosis of liver cells, and potentially to liver cancer [179]. Liver fibrosis on the other hand is a degenerative condition that spreads through normal liver tissue. Normal tissue gets replaced with increasing amounts of fibroblasts, which over time affects the functionality of the liver. This process is also linked to toxic substances which are not properly metabolized by the liver, eventually, this can also lead to the development of liver cancer [180].

The T stage represents the local growth of the tumor. T1 represents a small mass confined to the liver, and T4 represents an invasion of surrounding organs or blood vessels. The higher the T stage, the larger the tumor has grown or has invaded nearby organs or blood vessels. The higher the T stage, the worse the patient's prognosis [181]. Pathology (L) is the part of the TNM classification and means lymph invasion which indicates whether the tumor has metastasized to the lymph nodes. L0 indicates no lymph node metastasis and L1 indicates lymph node metastasis [182]. None of the 19 patients had lymph node metastasis. Pathology (G) represents the differentiation of tumor tissue. G1 represents the highest differentiation, the tumor tissue has the greatest similarity with normal tissue, the least malignancy, and a better prognosis. G3 represents poor differentiation, the tumor tissue has the greatest difference to normal tissue, the greatest malignancy, and the worst prognosis [183]. According to the TNM tumor assessment system, UICC divides the patient's status into I, II, III and IV. Stage I is an early-stage tumor, the tumor is relatively small, there is no metastasis, and the prognosis is relatively good after surgical intervention. Stage IV is an advanced tumor

stage, where the tumor has metastasized, and surgical intervention is no longer the best option. The prognosis is relatively poor in this stage [184].

T staging is an important part of the tumor staging system. T staging mainly describes the local situation of the primary site of the tumor [185]. It is generally understood that T staging is closely related to the prognosis of tumor patients. The higher the T staging score, the larger the tumor, the more invasive it is, and the worse the prognosis of the patient [186, 187]. For HCC, T3 or T4 means that the tumor has invaded the blood vessels, and the patient's prognosis is generally worse than that of low-stage patients [188].

In this study, we analyzed the PLVAP expression of 19 HCC patients with relevant clinical indicators and found that high expression of PLVAP in patient samples correlates with a lower T stage, while patients with lower PLVAP expression often had a higher T stage. This would indicate that PLVAP expression could be used to identify early stages of HCC due to high PLVAP expression. Whether PLVAP could also be targeted needs further studies investigating the molecular underpinnings of PLVAP-associated signaling in endothelial cells. Also, to consider is the currently low number of patients in this study, which needs to be addressed in future studies as well.

5 Conclusion

With the use of two different machine learning approaches on HCC sequencing data sets, we identified 7 genes that might be relevant as diagnostic or prognostic markers. Subsequent analysis revealed that not all of these genes were indeed useful as biomarkers or had a diagnostic value. However, one of the genes, PLVAP, showed good correlation as a biomarker as well as differential expression in tumor and normal liver tissue.

Further analysis predicted the expression of PLVAP in endothelial cells which was validated in HCC tumor and normal tissue. Here we focused on endothelial cells of the vascular system in the tumor. We found PLVAP expression in areas with vascular and

microvascular systems in tumors, whereas in normal liver tissue, almost no PLVAP expression was found.

Correlation of the HCC diagnosis of the patients analyzed in this study revealed an inverse correlation of PLVAP expression levels in endothelial cells to T staging. Our findings indicate that PLVAP has diagnostic and prognostic potential for HCC, as it appears to be regulated during tumor progression. High levels of PLVAP in early phases and a reduction in expression with increasing tumor size in later tumor stages were found. The data presented here provide an interesting observation that would warrant additional studies to better understand the relation between staging, PLVAP expression and angiogenesis signaling. In addition to PLVAP being a biomarker, it could also be targeted as part of a therapeutic intervention, given a better understanding of this interesting molecule.

References

1. Berkemeyer, S., *Primary Liver Cancers: Connecting the Dots of Cellular Studies and Epidemiology with Metabolomics.* Int J Mol Sci, 2023. **24**(3).
2. Sim, Y.K., et al., *Real-World Data on the Diagnosis, Treatment, and Management of Hepatocellular Carcinoma in the Asia-Pacific: The INSIGHT Study.* Liver Cancer, 2024. **13**(3): p. 298-313.
3. Tan, J.L., et al., *New High-Throughput Screening Identifies Compounds That Reduce Viability Specifically in Liver Cancer Cells That Express High Levels of SALL4 by Inhibiting Oxidative Phosphorylation.* Gastroenterology, 2019. **157**(6): p. 1615-1629.e17.
4. Sung, H., et al., *Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.* CA Cancer J Clin, 2021. **71**(3): p. 209-249.
5. Rumgay, H., et al., *Global burden of primary liver cancer in 2020 and predictions to 2040.* J Hepatol, 2022. **77**(6): p. 1598-1606.
6. Katanoda, K., Y. Ito, and T. Sobue, *International comparison of trends in cancer mortality: Japan has fallen behind in screening-related cancers.* Jpn J Clin Oncol, 2021. **51**(11): p. 1680-1686.
7. Yeo, Y., et al., *Viral hepatitis and liver cancer in Korea: an epidemiological perspective.* Asian Pac J Cancer Prev, 2013. **14**(11): p. 6227-31.
8. Petrick, J.L., et al., *International trends in liver cancer incidence, overall and by histologic subtype, 1978-2007.* Int J Cancer, 2016. **139**(7): p. 1534-45.
9. Paik, J.M., et al., *Changes in the Global Burden of Chronic Liver Diseases From 2012 to 2017: The Growing Impact of NAFLD.* Hepatology, 2020. **72**(5): p. 1605-1616.
10. Burton, A., et al., *Primary liver cancer in the UK: Incidence, incidence-based mortality, and survival by subtype, sex, and nation.* JHEP Rep, 2021. **3**(2): p. 100232.
11. Bencina, G., et al., *Burden and indirect cost of vaccine-preventable cancer mortality in Europe.* J Med Econ, 2024. **27**(sup2): p. 30-40.
12. Lefeuvre, C., H. Le Guillou-Guillemette, and A. Ducancelle, *A Pleiotropic Role of the Hepatitis B Virus Core Protein in Hepatocarcinogenesis.* Int J Mol Sci, 2021. **22**(24).
13. Fiehn, F., C. Beisel, and M. Binder, *Hepatitis C virus and hepatocellular carcinoma: carcinogenesis in the era of direct-acting antivirals.* Curr Opin Virol, 2024. **67**: p. 101423.
14. Bader, H., et al., *Association between Metabolic-Dysfunction-Associated Steatotic Liver Disease and Hepatic Cancer: Current Concepts and Future Challenges.* J Clin Med, 2024. **13**(11).
15. Capasso, M., et al., *The Role of Hepatitis Viruses as Drivers of Hepatocancerogenesis.*

Cancers (Basel), 2024. **16**(8).

16. Singal, A.G., P. Lampertico, and P. Nahon, *Epidemiology and surveillance for hepatocellular carcinoma: New trends.* J Hepatol, 2020. **72**(2): p. 250-261.
17. Zeng, R.W., et al., *Global prevalence, clinical characteristics, surveillance, treatment allocation, and outcomes of alcohol-associated hepatocellular carcinoma.* Clin Gastroenterol Hepatol, 2024.
18. Chen, P., et al., *Advances in the Pathogenesis of Metabolic Liver Disease-Related Hepatocellular Carcinoma.* J Hepatocell Carcinoma, 2024. **11**: p. 581-594.
19. Vallianou, N.G., et al., *NAFLD/MASLD and the Gut-Liver Axis: From Pathogenesis to Treatment Options.* Metabolites, 2024. **14**(7).
20. Samji, N.S., R. Verma, and S.K. Satapathy, *Magnitude of Nonalcoholic Fatty Liver Disease: Western Perspective.* J Clin Exp Hepatol, 2019. **9**(4): p. 497-505.
21. Dhanasekaran, R. and D.W. Felsher, *A Tale of Two Complications of Obesity: NASH and Hepatocellular Carcinoma.* Hepatology, 2019. **70**(3): p. 1056-1058.
22. Cassinotto, C., C. Aubé, and A. Dohan, *Diagnosis of hepatocellular carcinoma: An update on international guidelines.* Diagn Interv Imaging, 2017. **98**(5): p. 379-391.
23. Singal, A.G., M. Ng, and A. Kulkarni, *Advancing Surveillance Strategies for Hepatocellular Carcinoma: A New Era of Efficacy and Precision.* J Clin Exp Hepatol, 2024. **14**(6): p. 101448.
24. Murray, M.J., R.A. Huddart, and N. Coleman, *The present and future of serum diagnostic tests for testicular germ cell tumours.* Nat Rev Urol, 2016. **13**(12): p. 715-725.
25. Rastogi, A., *Changing role of histopathology in the diagnosis and management of hepatocellular carcinoma.* World J Gastroenterol, 2018. **24**(35): p. 4000-4013.
26. Dhanasekaran, R., A. Limaye, and R. Cabrera, *Hepatocellular carcinoma: current trends in worldwide epidemiology, risk factors, diagnosis, and therapeutics.* Hepat Med, 2012. **4**: p. 19-37.
27. Katayama, A., et al., *Atypia in breast pathology: what pathologists need to know.* Pathology, 2022. **54**(1): p. 20-31.
28. Menon, S.S., et al., *Ki-67 protein as a tumour proliferation marker.* Clin Chim Acta, 2019. **491**: p. 39-45.
29. Schreiber, L., et al., *Double immunolabeling by the RBM and the PLAP markers for identifying intratubular (in situ) germ cell neoplasia of the testis.* Int J Surg Pathol, 2003. **11**(1): p. 17-20.
30. Moradi, A., et al., *Beyond the biomarker role: prostate-specific antigen (PSA) in the prostate cancer microenvironment.* Cancer Metastasis Rev, 2019. **38**(3): p. 333-346.
31. Tümen, D., et al., *Pathogenesis and Current Treatment Strategies of Hepatocellular Carcinoma.* Biomedicines, 2022. **10**(12).

32. Shibata, T., *Genomic landscape of hepatocarcinogenesis.* J Hum Genet, 2021. **66**(9): p. 845-851.
33. Nagaraju, G.P., et al., *Epigenetics in hepatocellular carcinoma.* Semin Cancer Biol, 2022. **86**(Pt 3): p. 622-632.
34. Fu, S., J.D. Debes, and A. Boonstra, *DNA methylation markers in the detection of hepatocellular carcinoma.* Eur J Cancer, 2023. **191**: p. 112960.
35. Bardol, T., et al., *Circulating Tumor DNA Clinical Applications in Hepatocellular Carcinoma: Current Trends and Future Perspectives.* Clin Chem, 2024. **70**(1): p. 33-48.
36. Li, N., et al., *Contrast-enhanced ultrasound demonstrates greater adjuvant potential: past, current status, and future applications in Hepatocellular carcinoma early diagnosis.* Med Ultrason, 2023. **25**(3): p. 304-311.
37. Sharma, K.K., et al., *Diagnosis of the Initial Stage of Hepatocellular Carcinoma: A Review.* Curr Pharm Des, 2024. **30**(22): p. 1708-1724.
38. Outwater, E.K., *Imaging of the liver for hepatocellular cancer.* Cancer Control, 2010. **17**(2): p. 72-82.
39. Li, X.H., et al., *Diagnostic value of imaging examinations in patients with primary hepatocellular carcinoma.* World J Clin Cases, 2018. **6**(9): p. 242-248.
40. Hou, Y.C., et al., *Clinical whole-genome sequencing in cancer diagnosis.* Hum Mutat, 2022. **43**(11): p. 1519-1530.
41. Peeters, F., et al., *Applications of single-cell multi-omics in liver cancer.* JHEP Rep, 2024. **6**(7): p. 101094.
42. Safri, F., et al., *Heterogeneity of hepatocellular carcinoma: from mechanisms to clinical implications.* Cancer Gene Ther, 2024. **31**(8): p. 1105-1112.
43. Heimbach, J.K., et al., *AASLD guidelines for the treatment of hepatocellular carcinoma.* Hepatology, 2018. **67**(1): p. 358-380.
44. Kinsey, E. and H.M. Lee, *Management of Hepatocellular Carcinoma in 2024: The Multidisciplinary Paradigm in an Evolving Treatment Landscape.* Cancers (Basel), 2024. **16**(3).
45. Lindemann, J., J. Yu, and M.M.B. Doyle, *New horizons in liver transplantation for hepatocellular carcinoma.* S Afr J Surg, 2024. **62**(2): p. 8-12.
46. Li, L., et al., *Adjuvant Therapy for Hepatocellular Carcinoma After Curative Treatment: Several Unanswered Questions.* J Clin Transl Hepatol, 2024. **12**(5): p. 525-533.
47. Gerum, S., B. Grambozov, and F. Roeder, *Stereotactic body radiation therapy (SBRT) in patients with hepatocellular cancer-a narrative review and expert opinion.* J Gastrointest Oncol, 2024. **15**(4): p. 1880-1892.
48. Abualnil, A.Y., et al., *Role of Stereotactic Body Radiation Therapy in Hepatocellular Carcinoma.* Surg Oncol Clin N Am, 2024. **33**(1): p. 173-195.

49. Klink, A.J., et al., *Real-world Treatment Patterns and Reasons for Therapy Selection in Patients with Advanced Hepatocellular Carcinoma in US Oncology Practices.* Oncologist, 2022. **27**(3): p. e265-e272.
50. Ge, S. and D. Huang, *Systemic therapies for hepatocellular carcinoma.* Drug Discov Ther, 2015. **9**(5): p. 352-62.
51. Zhong, J., et al., *Conventional oral systemic chemotherapy for postoperative hepatocellular carcinoma: A systematic review.* Mol Clin Oncol, 2014. **2**(6): p. 1091-1096.
52. Chang, P.Y., et al., *Multidisciplinary Taiwan consensus for the use of conventional TACE in hepatocellular carcinoma treatment.* Front Oncol, 2023. **13**: p. 1186674.
53. Gao, F., et al., *Combination targeted therapy of VEGFR inhibitor, sorafenib, with an mTOR inhibitor, sirolimus induced a remakable response of rapid progressive Uterine PEComa.* Cancer Biol Ther, 2016. **17**(6): p. 595-8.
54. Zhang, M., et al., *Blockage of VEGF function by bevacizumab alleviates early-stage cerebrovascular dysfunction and improves cognitive function in a mouse model of Alzheimer's disease.* Transl Neurodegener, 2024. **13**(1): p. 1.
55. Shah, N.J., et al., *Product review on the Anti-PD-L1 antibody atezolizumab.* Hum Vaccin Immunother, 2018. **14**(2): p. 269-276.
56. Li, Z., et al., *Tumor Microenvironment Composition and Related Therapy in Hepatocellular Carcinoma.* J Hepatocell Carcinoma, 2023. **10**: p. 2083-2099.
57. Zubair, T. and D. Bandyopadhyay, *Small Molecule EGFR Inhibitors as Anti-Cancer Agents: Discovery, Mechanisms of Action, and Opportunities.* Int J Mol Sci, 2023. **24**(3).
58. Zhu, M., et al., *HER3 receptor and its role in the therapeutic management of metastatic breast cancer.* J Transl Med, 2024. **22**(1): p. 665.
59. Breuhahn, K. and P. Schirmacher, *Reactivation of the insulin-like growth factor-II signaling pathway in human hepatocellular carcinoma.* World J Gastroenterol, 2008. **14**(11): p. 1690-8.
60. Cabral, L.K.D., C. Tiribelli, and C.H.C. Sukowati, *Sorafenib Resistance in Hepatocellular Carcinoma: The Relevance of Genetic Heterogeneity.* Cancers (Basel), 2020. **12**(6).
61. Rimassa, L. and A. Santoro, *Sorafenib therapy in advanced hepatocellular carcinoma: the SHARP trial.* Expert Rev Anticancer Ther, 2009. **9**(6): p. 739-45.
62. Cheng, A.L., et al., *Efficacy and safety of sorafenib in patients with advanced hepatocellular carcinoma according to baseline status: subset analyses of the phase III Sorafenib Asia-Pacific trial.* Eur J Cancer, 2012. **48**(10): p. 1452-65.
63. Spratlin, J.L., et al., *Phase I pharmacologic and biologic study of ramucirumab (IMC-1121B), a fully human immunoglobulin G1 monoclonal antibody targeting the vascular endothelial growth factor receptor-2.* J Clin Oncol, 2010. **28**(5): p. 780-7.

64. Bantounou, M.A., et al., *Enfortumab vedotin and pembrolizumab as monotherapies and combination treatment in locally advanced or metastatic urothelial carcinoma: A narrative review.* Curr Urol, 2023. **17**(4): p. 271-279.
65. Finn, R.S., et al., *Pembrolizumab As Second-Line Therapy in Patients With Advanced Hepatocellular Carcinoma in KEYNOTE-240: A Randomized, Double-Blind, Phase III Trial.* J Clin Oncol, 2020. **38**(3): p. 193-202.
66. Krummel, M.F. and J.P. Allison, *CD28 and CTLA-4 have opposing effects on the response of T cells to stimulation.* J Exp Med, 1995. **182**(2): p. 459-65.
67. Waterhouse, P., et al., *Lymphoproliferative disorders with early lethality in mice deficient in Ctla-4.* Science, 1995. **270**(5238): p. 985-8.
68. Fife, B.T. and J.A. Bluestone, *Control of peripheral T-cell tolerance and autoimmunity via the CTLA-4 and PD-1 pathways.* Immunol Rev, 2008. **224**: p. 166-82.
69. Xiao, Y. and D. Yu, *Tumor microenvironment as a therapeutic target in cancer.* Pharmacol Ther, 2021. **221**: p. 107753.
70. Arneth, B., *Tumor Microenvironment.* Medicina (Kaunas), 2019. **56**(1).
71. Kalluri, R. and M. Zeisberg, *Fibroblasts in cancer.* Nat Rev Cancer, 2006. **6**(5): p. 392-401.
72. Coulson-Thomas, Y.M., et al., *The role of proteoglycans in the reactive stroma on tumor growth and progression.* Histol Histopathol, 2015. **30**(1): p. 33-41.
73. Riabov, V., et al., *Role of tumor associated macrophages in tumor angiogenesis and lymphangiogenesis.* Front Physiol, 2014. **5**: p. 75.
74. Reina-Campos, M., N.E. Scharping, and A.W. Goldrath, *CD8(+) T cell metabolism in infection and cancer.* Nat Rev Immunol, 2021. **21**(11): p. 718-738.
75. Malyshkina, A., et al., *Cytotoxic CD4(+) T cells in chronic viral infections and cancer.* Front Immunol, 2023. **14**: p. 1271236.
76. Shan, F., et al., *Therapeutic targeting of regulatory T cells in cancer.* Trends Cancer, 2022. **8**(11): p. 944-961.
77. Tay, C., A. Tanaka, and S. Sakaguchi, *Tumor-infiltrating regulatory T cells as targets of cancer immunotherapy.* Cancer Cell, 2023. **41**(3): p. 450-465.
78. Sharonov, G.V., et al., *B cells, plasma cells and antibody repertoires in the tumour microenvironment.* Nat Rev Immunol, 2020. **20**(5): p. 294-307.
79. Prager, I. and C. Watzl, *Mechanisms of natural killer cell-mediated cellular cytotoxicity.* J Leukoc Biol, 2019. **105**(6): p. 1319-1329.
80. Paolini, R., et al., *NK cells and interferons.* Cytokine Growth Factor Rev, 2015. **26**(2): p. 113-20.
81. Eelen, G., et al., *Basic and Therapeutic Aspects of Angiogenesis Updated.* Circ Res, 2020. **127**(2): p. 310-329.
82. Hida, K., et al., *Tumor angiogenesis--characteristics of tumor endothelial cells.* Int J

Clin Oncol, 2016. **21**(2): p. 206-212.

83. Llovet, J.M., et al., *Immunotherapies for hepatocellular carcinoma.* Nat Rev Clin Oncol, 2022. **19**(3): p. 151-172.
84. Kerbel, R.S., *Antiangiogenic therapy: a universal chemosensitization strategy for cancer?* Science, 2006. **312**(5777): p. 1171-5.
85. Fernández-Delgado, M., et al., *An extensive experimental survey of regression methods.* Neural Netw, 2019. **111**: p. 11-34.
86. Ali, H., et al., *Application and impact of Lasso regression in gastroenterology: A systematic review.* Indian J Gastroenterol, 2023. **42**(6): p. 780-790.
87. Song, Z., et al., *Identification of tryptophan metabolism-related genes in immunity and immunotherapy in Alzheimer's disease.* Aging (Albany NY), 2023. **15**(22): p. 13077-13099.
88. Popp, J., et al., *Markers of neuroinflammation associated with Alzheimer's disease pathology in older adults.* Brain Behav Immun, 2017. **62**: p. 203-211.
89. Shen, S., et al., *Identification of core genes and outcomes in hepatocellular carcinoma by bioinformatics analysis.* J Cell Biochem, 2019. **120**(6): p. 10069-10081.
90. Zhou, Y., et al., *Diagnosis and Prognosis of Non-small Cell Lung Cancer based on Machine Learning Algorithms.* Comb Chem High Throughput Screen, 2023. **26**(12): p. 2170-2183.
91. Guan, S., et al., *Identifying potential targets for preventing cancer progression through the PLA2G1B recombinant protein using bioinformatics and machine learning methods.* Int J Biol Macromol, 2024. **276**(Pt 1): p. 133918.
92. Fan, Y., et al., *COMPARE: classification of morphological patterns using adaptive regional elements.* IEEE Trans Med Imaging, 2007. **26**(1): p. 93-105.
93. Feng, Z., et al., *Machine learning-based quantitative texture analysis of CT images of small renal masses: Differentiation of angiomyolipoma without visible fat from renal cell carcinoma.* Eur Radiol, 2018. **28**(4): p. 1625-1633.
94. Stan, R.V., E. Tkachenko, and I.R. Niesman, *PV1 is a key structural component for the formation of the stomatal and fenestral diaphragms.* Mol Biol Cell, 2004. **15**(8): p. 3615-30.
95. Wen, Y., et al., *PLVAP protein expression correlated with microbial composition, clinicopathological features, and prognosis of patients with stomach adenocarcinoma.* J Cancer Res Clin Oncol, 2023. **149**(10): p. 7139-7153.
96. Wang, Y., et al., *Plasmalemma vesicle-associated protein promotes angiogenesis in cholangiocarcinoma via the DKK1/CKAP4/PI3K signaling pathway.* Oncogene, 2021. **40**(25): p. 4324-4337.
97. De Leo, L., et al., *Circulating PV-1 as a marker of celiac disease-associated liver injury.* Biomark Med, 2020. **14**(18): p. 1675-1681.

98. Wang, Y.H., et al., *Plasmalemmal Vesicle Associated Protein (PLVAP) as a therapeutic target for treatment of hepatocellular carcinoma.* BMC Cancer, 2014. **14**: p. 815.
99. *Integrated genomic analyses of ovarian carcinoma.* Nature, 2011. **474**(7353): p. 609-15.
100. Mingueneau, M., et al., *The transcriptional landscape of αβ T cell differentiation.* Nat Immunol, 2013. **14**(6): p. 619-32.
101. *Comprehensive genomic characterization defines human glioblastoma genes and core pathways.* Nature, 2008. **455**(7216): p. 1061-8.
102. Weinstein, J.N., et al., *The Cancer Genome Atlas Pan-Cancer analysis project.* Nat Genet, 2013. **45**(10): p. 1113-20.
103. *Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma.* Cell, 2017. **169**(7): p. 1327-1341.e23.
104. Shah, N.M., et al., *Pan-cancer analysis identifies tumor-specific antigens derived from transposable elements.* Nat Genet, 2023. **55**(4): p. 631-639.
105. Lin, C.P., et al., *Multimodal stimulation screens reveal unique and shared genes limiting T cell fitness.* Cancer Cell, 2024. **42**(4): p. 623-645.e10.
106. Xiao, H., et al., *Taurine Inhibits Ferroptosis Mediated by the Crosstalk between Tumor Cells and Tumor-Associated Macrophages in Prostate Cancer.* Adv Sci (Weinh), 2024. **11**(3): p. e2303894.
107. Liu, Y., et al., *Expanding PROTACtable genome universe of E3 ligases.* Nat Commun, 2023. **14**(1): p. 6509.
108. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
109. Dallaire, X. and C. Mérot, *A setback into a success: What can batch effects tell us about best practices in genomics?* Mol Ecol Resour, 2022. **22**(5): p. 1675-1677.
110. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-3.
111. Robin, X., et al., *pROC: an open-source package for R and S+ to analyze and compare ROC curves.* BMC Bioinformatics, 2011. **12**: p. 77.
112. Çorbacıoğlu Ş, K. and G. Aksel, *Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value.* Turk J Emerg Med, 2023. **23**(4): p. 195-198.
113. Tang, Z., et al., *GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses.* Nucleic Acids Res, 2017. **45**(W1): p. W98-w102.
114. DeLisser, H.M., P.J. Newman, and S.M. Albelda, *Platelet endothelial cell adhesion molecule (CD31).* Curr Top Microbiol Immunol, 1993. **184**: p. 37-45.
115. Pusztaszeri, M.P., W. Seelentag, and F.T. Bosman, *Immunohistochemical expression of endothelial markers CD31, CD34, von Willebrand factor, and Fli-1 in normal human*

tissues. J Histochem Cytochem, 2006. **54**(4): p. 385-95.

116. Dhakal, H.P., et al., *Vascularization in primary breast carcinomas: its prognostic significance and relationship with tumor cell dissemination.* Clin Cancer Res, 2008. **14**(8): p. 2341-50.
117. Ferrara, N., H.P. Gerber, and J. LeCouter, *The biology of VEGF and its receptors.* Nat Med, 2003. **9**(6): p. 669-76.
118. Li, Z., et al., *Anti-VEGFR2-interferon-α2 regulates the tumor microenvironment and exhibits potent antitumor efficacy against colorectal cancer.* Oncoimmunology, 2017. **6**(3): p. e1290038.
119. Mukaka, M.M., *Statistics corner: A guide to appropriate use of correlation coefficient in medical research.* Malawi Med J, 2012. **24**(3): p. 69-71.
120. Valenzuela, C.D., et al., *Characterization of adipose-derived mesenchymal stem cell combinations for vascularized bone engineering.* Tissue Eng Part A, 2013. **19**(11-12): p. 1373-85.
121. Pinto, E., et al., *Angiogenesis and Hepatocellular Carcinoma: From Molecular Mechanisms to Systemic Therapies.* Medicina (Kaunas), 2023. **59**(6).
122. Du, G., et al., *Regulatory T cells and immune escape in HCC: understanding the tumor microenvironment and advancing CAR-T cell therapy.* Front Immunol, 2024. **15**: p. 1431211.
123. Wu, X., et al., *Programmed cell death in hepatocellular carcinoma: mechanisms and therapeutic prospects.* Cell Death Discov, 2024. **10**(1): p. 356.
124. Liu, Y. and L. Liu, *Changes in the Epidemiology of Hepatocellular Carcinoma in Asia.* Cancers (Basel), 2022. **14**(18).
125. Tran, N.H., *Shifting Epidemiology of Hepatocellular Carcinoma in Far Eastern and Southeast Asian Patients: Explanations and Implications.* Curr Oncol Rep, 2022. **24**(2): p. 187-193.
126. Grgurevic, I., et al., *Hepatocellular Carcinoma in Non-Alcoholic Fatty Liver Disease: From Epidemiology to Diagnostic Approach.* Cancers (Basel), 2021. **13**(22).
127. Dhamija, E., S.B. Paul, and S. Kedia, *Non-alcoholic fatty liver disease associated with hepatocellular carcinoma: An increasing concern.* Indian J Med Res, 2019. **149**(1): p. 9-17.
128. Zhang, F., S. Ren, and Y. Zuo, *DC-SIGN, DC-SIGNR and LSECtin: C-type lectins for infection.* Int Rev Immunol, 2014. **33**(1): p. 54-66.
129. Yu, Q. and K. Gao, *CLEC4M overexpression inhibits progression and is associated with a favorable prognosis in hepatocellular carcinoma.* Mol Med Rep, 2020. **22**(3): p. 2245-2252.
130. Zhang, Y., et al., *CLEC4s as Potential Therapeutic Targets in Hepatocellular Carcinoma Microenvironment.* Front Cell Dev Biol, 2021. **9**: p. 681372.

131. Lal, G., et al., *Human Melanoma cells over-express extracellular matrix 1 (ECM1) which is regulated by TFAP2C.* PLoS One, 2013. **8**(9): p. e73953.
132. Xiong, G.P., et al., *Overexpression of ECM1 contributes to migration and invasion in cholangiocarcinoma cell.* Neoplasma, 2012. **59**(4): p. 409-15.
133. Lal, G., et al., *ECM1 expression in thyroid tumors--a comparison of real-time RT-PCR and IHC.* J Surg Res, 2008. **149**(1): p. 62-8.
134. Chen, H., W. Jia, and J. Li, *ECM1 promotes migration and invasion of hepatocellular carcinoma by inducing epithelial-mesenchymal transition.* World J Surg Oncol, 2016. **14**(1): p. 195.
135. Zhang, Z., et al., *VLM catecholaminergic neurons control tumor growth by regulating CD8(+) T cells.* Proc Natl Acad Sci U S A, 2021. **118**(28).
136. Block, I., et al., *CFP suppresses breast cancer cell growth by TES-mediated upregulation of the transcription factor DDIT3.* Oncogene, 2019. **38**(23): p. 4560-4573.
137. Farker, K., et al., *Impact of N-acetyltransferase polymorphism (NAT2) in hepatocellular carcinoma (HCC)--an investigation in a department of surgical medicine.* Exp Toxicol Pathol, 2003. **54**(5-6): p. 387-91.
138. Bi, J., et al., *CXCL14 inhibits the growth and promotes apoptosis of hepatocellular carcinoma cells via suppressing Akt/mTOR pathway.* J Recept Signal Transduct Res, 2021. **41**(6): p. 593-603.
139. Chew, S.C., S.Y. Choo, and P.K. Chow, *A new perspective on the immune escape mechanism in HCC: onco-foetal reprogramming.* Br J Cancer, 2021. **124**(12): p. 1897-1899.
140. Li, Z., et al., *Presence of onco-fetal neighborhoods in hepatocellular carcinoma is associated with relapse and response to immunotherapy.* Nat Cancer, 2024. **5**(1): p. 167-186.
141. Puntmann, V.O., *How-to guide on biomarkers: biomarker definitions, validation and applications with examples from cardiovascular disease.* Postgrad Med J, 2009. **85**(1008): p. 538-45.
142. Xi, S., et al., *Bioinformatics analysis to reveal key biomarkers for early and late hepatocellular carcinoma.* Transl Cancer Res, 2020. **9**(7): p. 4070-4079.
143. Vogel, C. and E.M. Marcotte, *Insights into the regulation of protein abundance from proteomic and transcriptomic analyses.* Nat Rev Genet, 2012. **13**(4): p. 227-32.
144. Yao, Y., et al., *Endoglin (CD105) expression in angiogenesis of primary hepatocellular carcinomas: analysis using tissue microarrays and comparisons with CD34 and VEGF.* Ann Clin Lab Sci, 2007. **37**(1): p. 39-48.
145. Ho, J.W., et al., *Clinicopathological and prognostic implications of endoglin (CD105) expression in hepatocellular carcinoma and its adjacent non-tumorous liver.* World J Gastroenterol, 2005. **11**(2): p. 176-81.

146. Steinberg, F., et al., *NIR reflection measurements of hemoglobin and cytochrome aa3 in healthy tissue and tumors. Correlations to oxygen consumption: preclinical and clinical data.* Adv Exp Med Biol, 1997. **428**: p. 69-77.
147. Leone, P., et al., *Endothelial cells in tumor microenvironment: insights and perspectives.* Front Immunol, 2024. **15**: p. 1367875.
148. Nagy, J.A., et al., *Why are tumour blood vessels abnormal and why is it important to know?* Br J Cancer, 2009. **100**(6): p. 865-9.
149. Baluk, P., H. Hashizume, and D.M. McDonald, *Cellular abnormalities of blood vessels as targets in cancer.* Curr Opin Genet Dev, 2005. **15**(1): p. 102-11.
150. Goltz, D. and H.P. Fischer, *[Liver sinusoids: pathology of endothelial findings].* Pathologe, 2020. **41**(5): p. 471-477.
151. Patten, D.A., et al., *Novel Targets in the Immune Microenvironment of the Hepatic Sinusoids for Treating Liver Diseases.* Semin Liver Dis, 2019. **39**(2): p. 111-123.
152. Brunt, E.M., et al., *Pathology of the liver sinusoids.* Histopathology, 2014. **64**(7): p. 907-20.
153. Vidal-Vanaclocha, F., et al., *Isolation and enrichment of two sublobular compartment-specific endothelial cell subpopulations from liver sinusoids.* Hepatology, 1993. **18**(2): p. 328-39.
154. DeLeve, L.D. and A.C. Maretti-Mira, *Liver Sinusoidal Endothelial Cell: An Update.* Semin Liver Dis, 2017. **37**(4): p. 377-387.
155. Bhandari, S., et al., *The Scavenger Function of Liver Sinusoidal Endothelial Cells in Health and Disease.* Front Physiol, 2021. **12**: p. 757469.
156. Wilkinson, A.L., M. Qurashi, and S. Shetty, *The Role of Sinusoidal Endothelial Cells in the Axis of Inflammation and Cancer Within the Liver.* Front Physiol, 2020. **11**: p. 990.
157. Feng, Y.X., et al., *Sinusoidal Endothelial Cell Progenitor Cells Promote Tumour Progression in Patients with Hepatocellular Carcinoma.* Stem Cells Int, 2020. **2020**: p. 8819523.
158. Bösmüller, H., et al., *Microvessel density and angiogenesis in primary hepatic malignancies: Differential expression of CD31 and VEGFR-2 in hepatocellular carcinoma and intrahepatic cholangiocarcinoma.* Pathol Res Pract, 2018. **214**(8): p. 1136-1141.
159. Segatelli, V., et al., *Evaluation and comparison of microvessel density using the markers CD34 and CD105 in regenerative nodules, dysplastic nodules and hepatocellular carcinoma.* Hepatol Int, 2014. **8**(2): p. 260-5.
160. Wu, L.Q., et al., *Can E-cadherin and CD34 be used as indicators of prognosis for hepatocellular carcinoma patients?* Clin Chem Lab Med, 2008. **46**(8): p. 1122-6.
161. Kasprzak, A. and A. Adamek, *Role of Endoglin (CD105) in the Progression of Hepatocellular Carcinoma and Anti-Angiogenic Therapy.* Int J Mol Sci, 2018. **19**(12).

162. Han, L., et al., *PBLD inhibits angiogenesis via impeding VEGF/VEGFR2-mediated microenvironmental cross-talk between HCC cells and endothelial cells.* Oncogene, 2022. **41**(13): p. 1851-1865.
163. Wang, Y., et al., *Integrin αVβ5/Akt/Sp1 pathway participates in matrix stiffness-mediated effects on VEGFR2 upregulation in vascular endothelial cells.* Am J Cancer Res, 2020. **10**(8): p. 2635-2648.
164. Lertkiatmongkol, P., et al., *Endothelial functions of platelet/endothelial cell adhesion molecule-1 (CD31).* Curr Opin Hematol, 2016. **23**(3): p. 253-9.
165. Figueiredo, C.C., et al., *Double immunofluorescence labeling for CD31 and CD105 as a marker for polyether polyurethane-induced angiogenesis in mice.* Histol Histopathol, 2019. **34**(3): p. 257-264.
166. Hassanpour, M., et al., *CD34 positive cells as endothelial progenitor cells in biology and medicine.* Front Cell Dev Biol, 2023. **11**: p. 1128134.
167. Nassiri, F., et al., *Endoglin (CD105): a review of its role in angiogenesis and tumor diagnosis, progression and therapy.* Anticancer Res, 2011. **31**(6): p. 2283-90.
168. Jeng, K.S., et al., *The Role of Endoglin in Hepatocellular Carcinoma.* Int J Mol Sci, 2021. **22**(6).
169. Rossi, E. and C. Bernabeu, *Novel vascular roles of human endoglin in pathophysiology.* J Thromb Haemost, 2023. **21**(9): p. 2327-2338.
170. Gao, F. and C. Yang, *Anti-VEGF/VEGFR2 Monoclonal Antibodies and their Combinations with PD-1/PD-L1 Inhibitors in Clinic.* Curr Cancer Drug Targets, 2020. **20**(1): p. 3-18.
171. Shah, A.A., M.A. Kamal, and S. Akhtar, *Tumor Angiogenesis and VEGFR-2: Mechanism, Pathways and Current Biological Therapeutic Interventions.* Curr Drug Metab, 2021. **22**(1): p. 50-59.
172. Lin, B., et al., *Anlotinib inhibits angiogenesis via suppressing the activation of VEGFR2, PDGFRβ and FGFR1.* Gene, 2018. **654**: p. 77-86.
173. Zhang, W., et al., *Clinicopathological features and prognostic factors of young patients with surgically treated liver cancer.* Medicine (Baltimore), 2015. **94**(12): p. e684.
174. Guo, H., et al., *Hepatocellular carcinoma in elderly: Clinical characteristics, treatments and outcomes compared with younger adults.* PLoS One, 2017. **12**(9): p. e0184160.
175. Federico, P., et al., *How to Treat Hepatocellular Carcinoma in Elderly Patients.* Pharmaceuticals (Basel), 2021. **14**(3).
176. Zhang, W., et al., *Differences in the prognostic value of tumor size on hepatocellular cancer-specific survival stratified by gender in a SEER population-based study.* United European Gastroenterol J, 2019. **7**(7): p. 933-941.
177. Llovet, J.M., et al., *Hepatocellular carcinoma.* Nat Rev Dis Primers, 2021. **7**(1): p. 6.
178. Naugler, W.E., et al., *Gender disparity in liver cancer due to sex differences in MyD88-*

dependent IL-6 production. Science, 2007. **317**(5834): p. 121-4.

179. Forner, A., J.M. Llovet, and J. Bruix, *Hepatocellular carcinoma.* Lancet, 2012. **379**(9822): p. 1245-55.
180. El-Serag, H.B., *Hepatocellular carcinoma.* N Engl J Med, 2011. **365**(12): p. 1118-27.
181. Li, Z., Q. Hong, and K. Li, *Nomogram predicting survival in patients with lymph node-negative hepatocellular carcinoma based on the SEER database and external validation.* Eur J Gastroenterol Hepatol, 2024. **36**(7): p. 904-915.
182. Li, S., Y. Shi, and Z. Wang, *The therapeutic value of ipsilateral supraclavicular lymphadenectomy in locally advanced breast cancer: A review.* Medicine (Baltimore), 2024. **103**(33): p. e39340.
183. Liu, D.Y., et al., *Analysis of lymph node metastasis and survival prognosis in early gastric cancer patients: A retrospective study.* World J Gastrointest Surg, 2024. **16**(6): p. 1637-1646.
184. Duseja, A., *Staging of hepatocellular carcinoma.* J Clin Exp Hepatol, 2014. **4**(Suppl 3): p. S74-9.
185. Edge, S.B. and C.C. Compton, *The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM.* Ann Surg Oncol, 2010. **17**(6): p. 1471-4.
186. Yao, N., et al., *Prognostic implications of T stage in different pathological types of colorectal cancer: an observational study using SEER population-based data.* BMJ Open, 2024. **14**(2): p. e076579.
187. Li, M., et al., *DOES T STAGE AFFECT PROGNOSIS IN PATIENTS WITH STAGE IV B DIFFERENTIATED THYROID CANCER?* Endocr Pract, 2019. **25**(9): p. 877-886.
188. Li, X., et al., *A nomogram to predict prognosis after surgery for young patients with hepatocellular carcinoma.* Transl Cancer Res, 2021. **10**(4): p. 1773-1786.

Acknowledgements

My time studying at the University of Munich is coming to an end. Looking back on my time in Munich, I have felt more care and kindness from the people around me during this fulfilling and often challenging, happy and often frustrating time. Here, I would like to especially thank my friends who have always helped and supported me.

Thanks to Prof. Werner. A few years ago, Prof. Werner gave me the precious opportunity to study at the University of Munich. Without this opportunity, I would not have been able to meet excellent professors, students and friends on this high-level platform in Munich. In my study and work, Prof. Werner also gave me help and advice to the best of his ability.

Thanks to Prof. Bazhin, Dr. Braun and Dr. Koch. In the process of completing the project, I often encountered some difficulties and problems. When I couldn't solve them, they gave me great help. They always gave me useful suggestions and patiently answered my questions. It can be said that without their support, this project would not be completed.

I would like to thank the students in the laboratory. When my experiment was not going well, they comforted me and encouraged me to overcome difficulties and persevere to the end. Here I need to mention Liu Yunfei in particular. When I first arrived at the laboratory and was not familiar with the laboratory, he enthusiastically helped me get familiar with the environment and gave me unreserved guidance on experimental techniques.

Finally, I want to thank my family. They have been silently supporting and encouraging me. With their care, I can devote myself to study and work.

Thanks again to everyone for their help, care and support, and I wish you all a smooth work and life and good health in the future.

Affidavit

Affidavit

Li, Zhongyi
Surname, first name

I hereby declare, that the submitted thesis entitled:

" *Clinical value of Plasmalemma Vesicle Associated Protein in Hepatocellular Carcinoma"*

is my own work. I have only used the sources indicated and have not made unauthorized use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.
I further declare that the dissertation presented here has not been submitted in the same or similar form to any other institution for the purpose of obtaining an academic degree.

Hubei 24.06.2025 Li Zhongyi

place, date Signature doctoral candidate

Confirmation of congruency between printed and electronic version of the doctoral thesis

Doctoral candidate: Zhongyi Li

I hereby declare that the electronic version of the submitted thesis, entitled

" Clinical value of Plasmalemma Vesicle Associated Protein in Hepatocellular Carcinoma"

is congruent with the printed version both in content and format.

Hubei 24.06.2025 Li Zhongyi

place, date Signature doctoral candidate