eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Departamento de Fisiología
Facultad de Medicina y Enfermería

Cholangiocarcinoma biomarkers in biofluid extracellular vesicles: a new liquid biopsy strategy

Tesis presentada por

AINHOA LAPITZ DAMBOLENEA

Donostia – San Sebastián

2022

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

biodonostia
osasun ikerketa institutua
instituto de investigación sanitaria

Cholangiocarcinoma biomarkers in biofluid extracellular vesicles: a new liquid biopsy strategy

Tesis presentada por

Ainhoa Lapitz Dambolenea

Para la obtención del título de doctora en

***Investigación Biomédica* por la**

Universidad del País Vasco/Euskal Herriko Unibertsitatea

Tesis dirigida por

Dr. D. Jesús María Bañales Asurmendi

Dr. D. Pedro Miguel Rodrigues

The project exposed in this thesis was performed at Biodonostia Health Research Institute under the supervision of Dr. Jesús Mª Bañales Asurmendi and Dr. Pedro Miguel Rodrigues.

Ainhoa Lapitz Dambolenea was the recipient of a Ph.D. fellowship (PRE_2017_1_0345) from Gobierno Vasco (Basque Country, Spain) and a Short-Term fellowship (Copenhagen, Denmark: *Reference number: EP_2020_1_0030*) from Gobierno Vasco (Basque Country, Spain).

This work was funded by Spanish Carlos III Health Institute (ISCIII) [JMB (Miguel Servet Program CON14/00129 and CPII19/00008) PMR (Sara Borrell CD19/00254)] cofinanced by "Fondo Europeo de Desarrollo Regional" (FEDER); CIBERehd (ISCIII); IKERBASQUE, Basque foundation for Science (to JMB and PMR); "Diputación Foral de Gipuzkoa" (2020-CIEN-000067-01 to PMR); Department of Health of the Basque Country (2017111010, 2020111077 to JMB), BIOEF (Basque Foundation for Innovation and Health Research: EiTB Maratoia BIO15/CA/016/BD to JMB). "Fundación Científica de la Asociación Española Contra el Cáncer" (AECC Scientific Foundation, to JMB). PSC Partners US (to JMB) and PSC Supports UK (to JMB: 06119JB). European Union's Horizon 2020 Research and Innovation Program (grant number 825510, ESCALON: to JMB). AMMF-The Cholangiocarcinoma Charity (EU/2019/AMMFt/001, to JMB and PMR). The funding sources had no involvement in the study design, data collection and analysis, decision to publish, or preparation of these articles.

TABLE OF CONTENTS

PART I

GENERAL INTRODUCTION

CHAPTER 1

CHOLANGIOCARCINOMA

1. GENERAL FEATURES

Cholangiocarcinomas (CCAs) comprise a heterogeneous group of malignancies displaying features of biliary differentiation that can arise at any point along the biliary tree, from the canals of Hering to the main bile duct.[1–3] According to the origin of these tumors, CCAs may emerge mainly from the malignant transformation of the epithelial cells lining the bile ducts (*i.e.*, cholangiocytes), though they can also be generated from hepatic stem cells, progenitor cells in peribiliary glands or even from hepatocytes undergoing trans-differentiation.[4–6]

CCAs are highly desmoplastic tumors, with extensive stroma. This tumor microenvironment (TME), mostly composed of cancer-associated endothelial cells, cancer-associated fibroblasts, a complex group of inflammatory cells including macrophages, neutrophils, natural killer cells and T cells together with the acellular fraction termed as extracellular matrix, supports the epithelial proliferation of tumor cells consequently fuelling tumor growth.[3,7]

2. CLASSIFICATION

2.1 Anatomical classification

According to the primary anatomical site of origin, the World Health Organization (WHO) newest International Classification of Disease (ICD) coding system version ICD-11 classifies CCAs as intrahepatic CCA (iCCA), perihilar CCA (pCCA) or distal CCA (dCCA) (**Figure 1.1**).[3,8] This new classification assigns individual ICD codes for the three different subtypes: 2C12.10 for iCCA, 2C18.0 for pCCA, and 2C15.0 for dCCA, and will be effective from 1st of January of 2022.

iCCAs arise at any point of the intrahepatic biliary tree, emerging from the smallest branches called bile ductules to the second-order bile ducts also known as segmental bile ducts. iCCAs represent the least common subtype, accounting for 10-20% of all CCA malignancies. pCCAs, formerly known as Klatskin tumors after their description by Klatskin in 1965, occur between the second order right and/or left hepatic bile ducts and the insertion of the cystic duct into the common bile duct.[9] pCCAs are the most usual subtype of CCAs, as they comprise 50 to 60% of all cases. Lastly, dCCAs are found in the common bile duct below the cystic duct through the ampulla of Vater, where the bile

duct connects with the pancreatic duct, and they account for 20-30% of all bile duct cancers.

Each anatomic subtype is associated to different risk factors, genetic aberrations, growing patterns, clinical presentations, diagnostic strategies, therapeutic options, prognosis, and clinical management approaches, thus representing independent entities. Therefore, the term extrahepatic cholangiocarcinomas (eCCAs) which grouped until very recently pCCAs and dCCAs is now strongly discouraged.

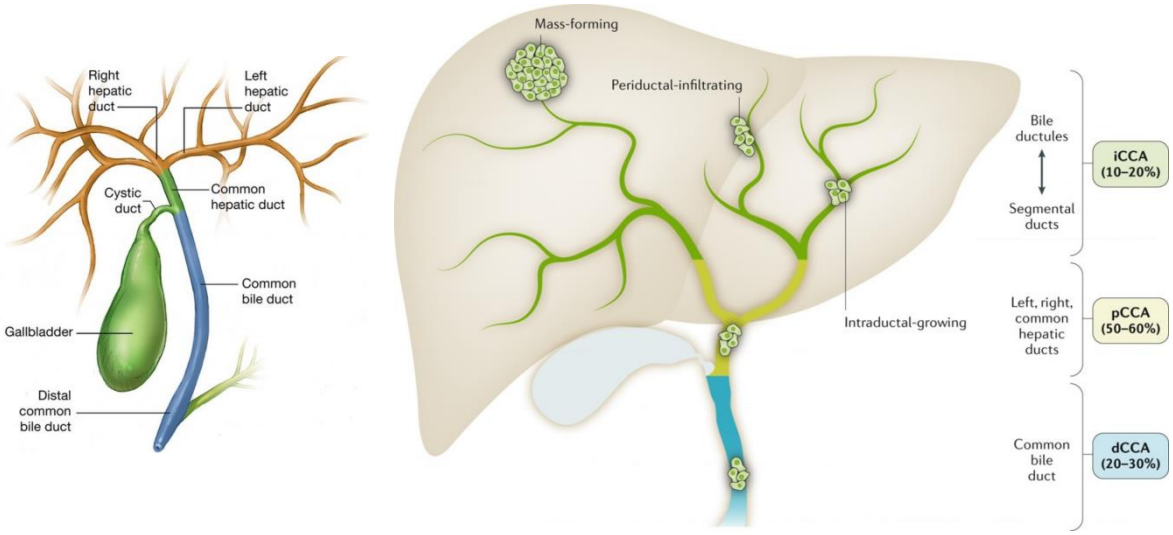


Figure 1.1. Bile duct anatomy (left, © 2005-2011 American Society of Clinical Oncology (ASCO)**) and anatomical classification of cholangiocarcinomas (right)**.[3] Depending on the site of origin, CCAs are classified as intrahepatic (iCCA), perihilar (pCCA) or distal (dCCA).

2.2 Tumor staging

CCA tumors are generally staged according to the American Joint of Committee on Cancer (AJCC) guidelines. The current version of the staging manual is the Eight Edition, effective since 2018, which classifies CCA tumors in consonance with the tumor node and metastasis (TNM) system.[10] TNM abbreviation refers to: (T), size of the primary main tumor; (N), spread of cancer to regional lymph nodes; and (M), dissemination of cancer to distant parts of the body or metastasis. Stages rage from I to IV; the lower the number, the less the cancer has spread. In line with CCA heterogeneity, different staging definitions apply for intrahepatic, perihilar and distal bile duct tumors (**Table 1.1**).

Table 1: CCA stages according to the 8th Edition of the AJCC staging manual.[10]

	AJCC Stage	Stage grouping		
iCCA	0	Tis	N0	M0
	IA	T1a	N0	M0
	IB	T1b	N0	M0
	II	T2	N0	M0
	IIIA	T3	N0	M0
	IIIB	T4	N0	M0
		AnyT	N1	M0
	IV	AnyT	AnyN	M1
pCCA	0	Tis	N0	M0
	I	T1	N0	M0
	II	T2a or T2b	N0	M0
	IIIA	T3	N0	M0
	IIIB	T4	N0	M0
	IIIC	Any T	N1	M0
	IVA	Any T	N2	M0
	IVB	Any T	Any N	M1
dCCA	0	Tis	N0	M0
	I	T1	N0	M0
	IIA	T2	N0	M0
		T1	N1	M0
	IIB	T3	N0	M0
		T2 or T3	N1	M0
	IIIA	T1, T2 or T3	N2	M0
	IIIB	T4	Any N	M0
	IV	Any T	Any N	M1

3. EPIDEMIOLOGY

CCA represents the 2nd most common primary liver cancer after hepatocellular carcinoma (HCC), accounting for 10-15% of all hepatic malignancies and 3% of all gastrointestinal cancers.[3]

It generally arises in people in their sixties-seventies, although younger individuals may also be affected. According to the biological sex, CCA occurs in both male and females, but a slightly higher incidence has been reported in males.[3]

3.1 Incidence

In most countries, CCA is considered a rare disease, as the recorded cases are lower than 6 per 100,000 people. However, its incidence is increasing globally, representing a major health and social problem.[1] The worldwide incidence of CCA is heterogeneous,

being higher in Eastern countries compared to the Western ones, potentially reflecting geographical differences in the prevalence of CCA-related risk factors.[11] The highest CCA age-standardized incidence was recorded in Northeast Thailand, South Korea and in Shanghai (China), with rates of 85, 7.1-8.8 and 7.55 per 100,000 people, respectively.[1,11,12] In contrast, in most Western countries it ranges from 2 to 6 cases per 100,000 people per year.

Over the past few decades, the incidence rates of CCA appear to have changed and different subtypes of CCA seem to show distinct epidemiological trends. Latest studies reported increasing incidence in iCCA and decreasing in eCCA implying the potential growing trend of iCCA identification to advances in diagnostic techniques and changes in some related risk factors.[12] Still, plausible explanations behind the trends in CCA incidence are complex and reported changes in incidence rates need to be interpreted with caution. Indeed, the previous coding system ICD-10 lacked a separated code for pCCA, the commonest form of CCA, which could have been incorrectly coded in patient registries as either iCCA or dCCA in the past decades.[13]

3.2 Mortality

The silent growth of CCAs leads to a late diagnosis, which combined with their highly aggressive nature, chemoresistance and the limited available therapeutic options available markedly contribute to a high CCA-associated mortality rate. Thus, CCA is responsible of 2% of all cancer-related deaths worldwide.[3,14]

The median overall survival of patients with CCA is approximately 6 months, with a 5-year survival of only 7-20% of individuals.[3] Paralleling the incidence trends, the worldwide CCA-related mortality is higher in men than in women and it is also rising annually.

The annual mortality rates due to CCA are heterogeneously distributed globally, pointing out Latin America, Lithuania and Czech Republic as the countries/regions with the lowest (<2 deaths per 100,000 people) and Japan, Hong Kong and Austria with the highest (>4 deaths per 100,000 people) mortality rates **(Figure 1.2)**.[3,14] CCA mortality in countries/regions with the highest incidence rates including South Korea, Taiwan, China and Thailand has not been studied yet and future studies on this topic are eagerly awaited.

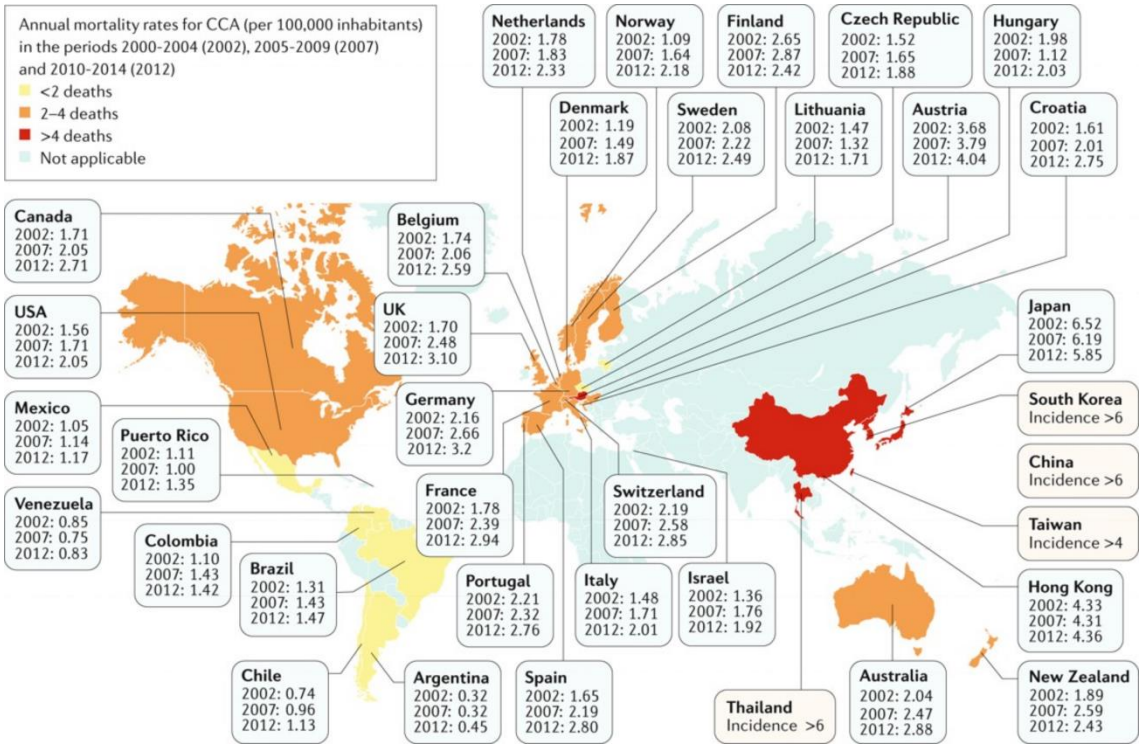


Figure 1.2. A**ge-standardized annual mortality rates of CCA worldwide**.[3] Mortality rates including iCCA, pCCA and dCCA are reported, from 2000 to 2014. Yellow-filled countries/regions indicate low mortality (<2 deaths per 100,000 people), orange-filled countries/regions indicate mortality between 2 and 4 deaths per 100,000 people, and read-filled countries/regions indicate high mortality (>4 deaths per 100,000 people). Incidence is displayed in highly prevalent CCA regions where mortality rates were not reported.

4. RISK FACTORS

The etiology of most bile duct cancers remains elusive, considering them as sporadic. In this sense, in Western countries, around 50% of cases are being diagnosed without any referable predisposing condition.

Still, epidemiological studies have suggested that there are also some well-established risk factors that may increase the odds of cholangiocarcinogenesis, part of them related to specific CCA subtypes while others common to all CCAs. Different degrees of predisposition have been assigned to the established risk factors, some of them accounting for a high risk of CCA development but with a low population frequency, while others being more common but with a milder association with this malignancy (**Figure 1.3**).[11] Additionally, there is also a spatial-temporal segregation of the underlying etiological factors of CCA, with some of them varying according to different geographical and historical aspects.

4.1 Bile duct disorders

4.1.1 Choledochal cysts

Choledochal cysts are congenital disorders of the biliary tree characterized by cystic dilatation of the intrahepatic and/or extrahepatic biliary tree.[15] It is a rare inherited disease more common in East communities rather than Western populations, with incidence rates ranging from 1 in 1000 in some Asian populations and from 1 in 100000-150000 in Western inhabitants. Patients with choledochal cysts may develop CCA at a mean age of 30 years, much younger than the general population, and a recent meta-analysis has reported an association between bile duct cysts and CCA development with odds ratios (ORs) of 26.71 for iCCA and 34.94 for eCCA.[16]

4.1.2 Hepatolithiasis, choledocholithiasis, cholelithiasis and cholecystolithiasis

Hepatolithiasis refers to the presence of calculi in the intrahepatic biliary tree. This bile duct disorder is quite frequent in East Asia, while it is rather unusual in Western Countries.[17] In consonance, iCCA has been detected in 5-13% of patients with hepatolithiasis.[17]

Regarding choledocholithiasis, cholelithiasis and cholecystolithiasis, both choledocholithiasis and cholelithiasis have been strongly associated with iCCA and eCCA in a recent meta-analysis (choledocholithiasis, OR 10.08 iCCA and 18.58 eCCA; cholelithiasis, OR 3.38 iCCA and 5.92 eCCA), while cholecystolithiasis was only significantly associated to the development of eCCAs (OR=2.94).[16]

4.1.3 Primary sclerosing cholangitis

Primary sclerosing cholangitis (PSC) is a chronic cholestatic and immune-mediated disease of unknown etiology, characterized by the development of multifocal fibro-inflammatory biliary strictures, leading to the subsequent obstruction of intra- and/or extrahepatic bile ducts.[18]

PSC-related CCA has a geographical distribution that follows the incidence of PSC, with an observed ascending gradient from the Eastern to the Western and from the Southern to the Northern countries.[19] The association between CCA (mainly pCCA) and PSC is one of the most reported one, especially in Europe. The lifetime risk of developing CCA in the setting of PSC is 400 times higher compared to the general population, with 10-

20% of patients with PSC developing CCA throughout their life.[20] CCA in a background of PSC is usually diagnosed in the fourth decade of life, contrasting the reported seventh decade in general population. Furthermore, up to 50% of tumors are detected within the first year after PSC diagnosis. A population-based study of the Surveillance, Epidemiology, and End Results (SEER)-Medicare US registry reported strong association between PSC and CCA development, where PSC appeared to increase the odds for iCCA by ~22-fold and ~41-fold for eCCA.[21] In agreement, the most recent meta-analysis including 11 cohort studies revealed a strong positive association between PSC and the risk of cholangiocarcinogenesis, with a relative risk of 584.37 for CCA development in patients with PSC, compared to individuals without PSC.[22]

4.1.4 Caroli disease

Caroli's syndrome is a rare autosomal recessive congenital condition characterized by non-obstructive saccular or fusiform dilatation of larger segmental intrahepatic bile ducts.[23]

This congenital biliary tract malformation also predisposes to the development of CCA. Indeed, it is one of the strongest risk factor for both iCCA and eCCA, conferring a 38-fold higher risk of iCCA and a 97-fold higher risk of eCCA.[21]

4.2 Parasitic infections

Infections with *Opistorchis viverrini* and *Clonorchis sinensis* trematodes (flatworm parasites, commonly called flukes) are a major cause of CCA in Southeast Asian regions including Korea, China, Taiwan, Vietnam and far Eastern Russia.[2,24] *O. viverrini* and *C. sinensis* are food-borne trematode parasites that may infect general population after the ingestion of contaminated raw and/or undercooked fish. These parasites allocate into the bile ducts and feed on bile, thus promoting biliary injury and inflammation.

In Southeast Asia, chronic infections with liver flukes have been causally related to CCA, being the vast majority of CCAs linked to these parasitic infestations in the mentioned endemic areas.[25] In East Asia, up to 10% of inhabitants chronically infected with *O. viverrini* and *C. sinensis* are known to develop CCA, especially iCCA, where iCCA represents around 85% of all primary liver cancers in that area. A meta-analysis of case-control studies reported a 4.8 overall relative risk of CCA development in patients infected with either *C. sinensis* or *O. viverrini* liver flukes.[26]

4.3 Viral infections

Chronic infections with hepatitis B (HBV) or C (HCV) viruses represent risk factors for CCA, with a stronger association for iCCA. According to a meta-analysis including 25 case-control studies, an OR of 4.57 and 4.28 was reported for HBV and HCV in iCCA, respectively, while a milder association to eCCA was observed (OR HBV = 2.11 and OR HCV = 1.98).[16]

The precise mechanisms by which HBV and HCV directly cause cancer (rather than the associated cirrhosis *per se*) are not completely deciphered yet, but some murine studies have suggested that hepatitis viruses can infect hepatic progenitor cells (HPC) and induce their expansion and transformation, leading to increased risk of liver tumor development.[27]

4.4 Liver diseases

4.4.1 Cirrhosis

Cirrhosis, a well-established risk factor for HCC, has also been described to increase the odds of CCA development. A meta-analysis including fourteen case-control studies with iCCA and eight case-control studies with eCCA identified cirrhosis as a strong risk factor for iCCA (OR 15.32) and in a lower extent for eCCA (OR 3.82).[16]

The raised risk of CCA, particularly iCCA, in patients with cirrhosis is explained by the chronic hepatic injury, inflammation, cell death, and subsequent occurrence of fibrosis and regenerative responses.[11]

4.4.2 Hemochromatosis

Hereditary hemochromatosis is an inherited condition resulting in a deregulated iron absorption that can lead to total body iron overload with secondary tissue damage in a wide range of organs, including the liver.[28] Results from the US SEER registry reported a 2.07-fold increased risk of iCCA development in patients with hemochromatosis, whereas no association was found for eCCA.[21]

4.5 Gastrointestinal diseases

4.5.1 Inflammatory bowel disease

Inflammatory bowel disease (IBD) is a term gathering two well-known conditions, Crohn's disease and ulcerative colitis (UC), that are characterized by chronic inflammation of the gastrointestinal tract. It has been suggested that inflammatory conditions of the digestive tract can cause bile duct inflammation and several studies have reported an increased risk of CCA development in patients with IBD, with a stronger association detected for patients with UC.[11] Nonetheless, as 70-80% of patients with PSC have concomitant UC, the association of IBD with CCA may depend on the presence of PSC, so the direct impact of IBD in increasing CCA risk still needs to be proved.[11]

4.5.2 Chronic pancreatitis

Chronic pancreatitis is a multifactorial, fibro-inflammatory syndrome in which repetitive episodes of pancreatic inflammation lead to extensive fibrotic tissue replacement, resulting in chronic pain together with exocrine and endocrine pancreatic insufficiency.[29]

A positive association between chronic pancreatitis and CCA has been reported, with a stronger association for eCCA (OR=6.61) than iCCA (OR=2.66).[21]

4.6 Metabolic and endocrine disorders

The role of metabolic and endocrine disorders including type II diabetes, NAFLD/NASH and obesity in increasing risk of CCA as single etiologic factors has been described in some studies. Nevertheless, these syndromes usually co-occur and the relative collective impact of these overlapping diseases in increasing the risk of bile duct cancer development still remains to be elucidated.

4.6.1 Type II diabetes *mellitus*

A meta-analysis conducted in 2012 evaluated the association between type II diabetes and the risk of developing CCA, reporting a higher risk of CCA development in patients with diabetes compared to individuals without diabetes (iCCA RR=1.97 and eCCA RR=1.63).[30] Additionally, a positive association between type II diabetes and CCA was

also reported in the SEER-Medicare registry, with an OR of 1.54 and 1.45 for iCCA and eCCA, respectively, with no differences among intrahepatic and extrahepatic subtypes.[21]

4.6.2 NAFLD / NASH

Non-alcoholic fatty liver disease (NAFLD) is the most prevalent cause of chronic liver disease in Western countries.[31] Nearly 30-40% of individuals with NAFLD develop non-alcoholic steatohepatitis (NASH), and among patients with NASH hepatic fibrosis occurs in 40-50% of cases.

A meta-analysis carried out in 2017 indicated that NAFLD increases the odds for CCA development, particularly for iCCA (*i.e.*, iCCA OR=2.2 and eCCA OR=1.5).[32]

4.6.3 Obesity

Obesity is a major public health problem worldwide, whose prevalence is exponentially rising, especially in Western countries.

Several epidemiological studies have shown that obesity may be positively correlated with CCA occurrence, principally in western regions, while most studies including the Asian population reported a null effect of obesity on CCA development.[33] Different meta-analysis have also revealed controversial results,[34–36] and the most recent one published in 2020 including both Asian and Western countries did not found any statistically significant association with iCCA neither with eCCA.[16]

Therefore, the potential risk of CCA development in obese patients remains controversial, but owing to the fact that obesity is becoming a global pandemic, the potential obesity-associated CCA hazard *per se* deserves future central attention.

4.7 Life-style

4.7.1 High alcohol consumption

A meta-analysis including fifteen case-control studies of patients with iCCA as well as eleven case-control studies of patients with eCCA revealed a positive association between alcohol consumption and the development of both iCCA and eCCA, with OR values of 3.15 and 1.75, respectively.[16]

4.7.2 Tobacco smoking

Tobacco smoking has also been tested as a potential risk for CCA development. A recent meta-analysis involving 324,333 smoker and non-smoker participants revealed a moderate association between tobacco smoking and risk of CCA (OR 1.31). Both iCCA and eCCA were associated with an increased risk of CCA development, with ORs of 1.31 and 1.32, respectively.[37]

4.8 Environmental exposure

Exposure to several toxic and environmental factors is known or suspected to be related to CCA development. Among them, Thorotrast, 1,2-Dichloropropane, asbestos, nitrosamines, dioxines and vinyl chlorides have been previously linked to cholangiocarcinogenesis.

4.8.1 Thorotrast

Thorotrast is a suspension containing particles of the radioactive compound thorium dioxide (ThO_2) that was administered as a radiocontrast agent in medical radiography from 1930 to 1950.[38] Individuals have developed CCA decades after Thorotrast administration and two studies have reported a 303-fold increased CCA risk in subjects exposed to this radiologic contrast agent compared to the general population in Japan.[39,40]

4.8.2 1,2-Dichloropropane

A retrospective study, which examined the CCA incidence in employees of a small Japanese printing company in Osaka who had been exposed to 1,2-Dichloropropane between 1987 and 2006, demonstrated the implication of chronic exposure to 1,2-Dichloropropane, an organic solvent classified as chlorocarbon traditionally used in printing, as a causative effect of CCA development, with an incidence rate ratio of 14.90.[41]

4.8.3 Asbestos

Several case-control studies also suggested an increased risk of liver cancer development in subjects exposed to asbestos. Whilst the evidence linking the exposure to this fibrous silicate mineral to eCCA was inconclusive, the association with intrahepatic bile duct malignancies seems robust.[42,43]

4.9 Genetic polymorphisms

Preliminary evidences have supported a potential association between CCA and polymorphisms in genes encoding enzymes involved in xenobiotic detoxification [*i.e.*, glutathione S transferases *GSTM1* and *GSTT1,* aryl-hydrocarbon hydroxylase coded by *CYP1A1*], DNA repair [*i.e.*, human oxoguanine glycosylase 1 *hOGG1*, MutY homolog *MUTYH*], immune response [*i.e.*, natural killer cell receptor G2D *NKG2D*], carcinogenesis [*i.e.*, multidrug resistance-associated protein 2 *MRP2/ABCC2*] and folate metabolism [*i.e.*, 5,10-methylenetetrahydrofolate reductase *MTHFR*], among others.[11]

These initial studies were not strong enough to support a clear relationship between specific polymorphisms and tumor development, but the fact that most CCA cases cannot be explained by the currently identifiable and established risk factors may indicate the existence of significant genetic components responsible for CCA pathogenesis. This fact currently requires a clear identification by robust whole genome sequencing techniques. In this sense, recently, the first genome wide association study (GWAS) of CCA has been approved (PLCO-631) which aims to identify genetic risk single nucleotide polymorphisms (SNP) variants associated with CCA development in a large, well-powered GWAS study (https://cdas.cancer.gov/approved-projects/2627). This international multicentre study is currently recruiting patient samples.

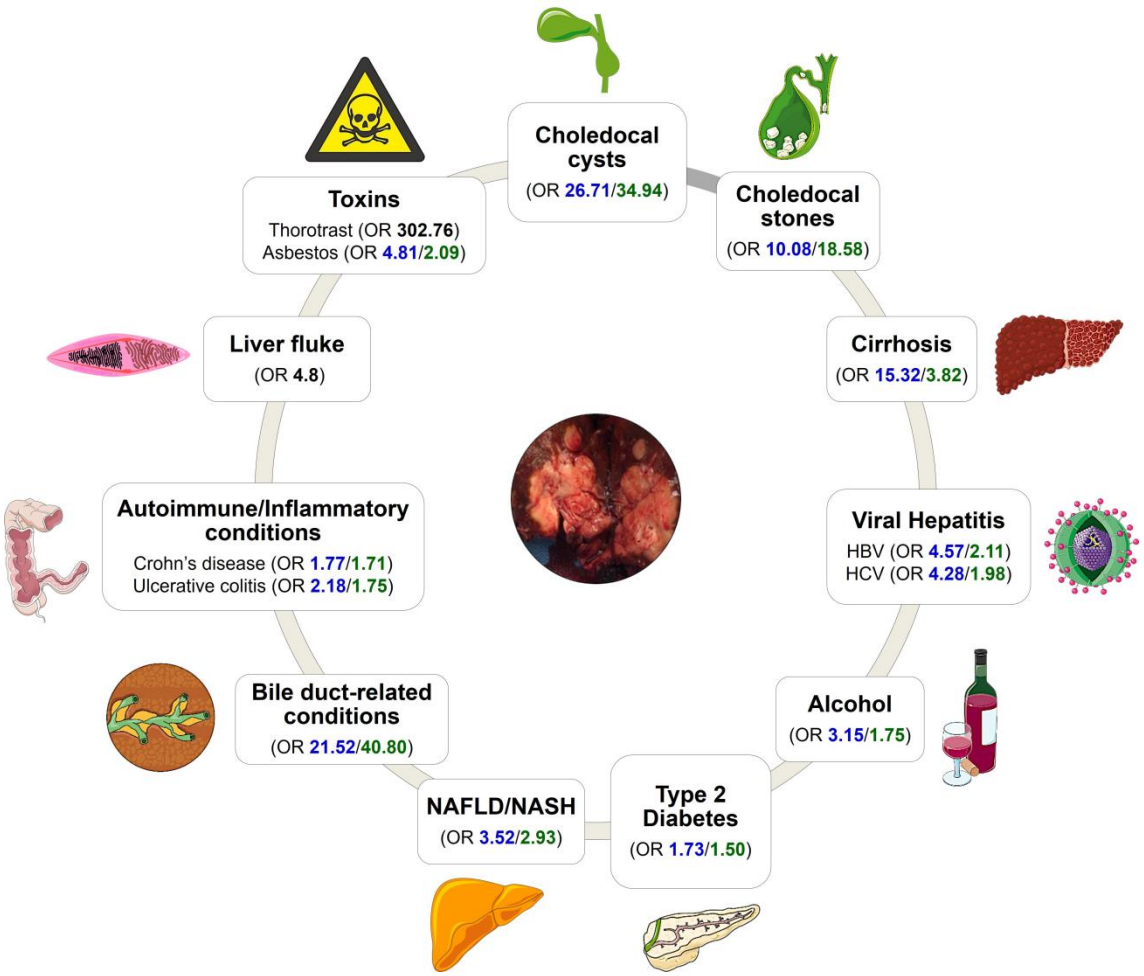


Figure 1.3: Risk factors for iCCA and eCCA. Blue OR are assigned to iCCAs while green ORs refer to eCCAs. Black ORs report the risk for CCA regardless the anatomical origin of the tumor. Abbreviations: HBV, hepatitis B virus; HCV, hepatitis C virus; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; OR, odds ratio.[44]

5. DIAGNOSIS

CCAs are generally asymptomatic at early stages, resulting in an advanced-stage diagnosis in most patients (~70%), when the disease is already widespread, limiting the current therapeutic options and resulting in dismal prognosis.[45]

Although there are no specific symptoms, some clinical manifestations might appear in patients with bile duct malignancy. For instance, jaundice is the most common symptom in pCCA and dCCA diseases due to biliary tract obstruction and subsequent obstructive cholestasis.[1] Nevertheless, jaundice is also a common symptom for many other medical conditions such as choledocholithiasis, parasitic infections, PSC, biliary cysts, duodenal diverticula, haemobilia, pancreatitis and malignancies including pancreatic and ampullary neoplasies.[45] Patients with iCCA, due to their often late presentation, are more

likely to present with even more nonspecific symptoms than jaundice such as fever, malaise, unexplained weight loss, fatigue, pruritus and/or abdominal pain. Accordingly, iCCA is an incidental finding in around 20-25% of cases.[46]

Currently, diagnosis of CCA requires a combination of clinical, radiological and/or nonspecific biochemical markers, with tumor biopsy still remaining as the only conclusive way to confirm cancer presence.

5.1 Blood biochemical tests and serum tumor markers

An increase in serum levels of biliary tract-excreted products such as bilirubin, alkaline phosphatase (ALP), gamma-glutamyl transpeptidase (GGT), 5'-nucleosidase or cholesterol are quite common in patients with CCA, generally due to obstructive cholestasis, which makes bile turn back into the bloodstream, yellowing the skin and resulting in pale stools. On the other hand, the most common liver damage markers alanine transaminase (ALT) and aspartate transaminase (AST) may be normal, particularly at early tumor stages and in the absence of cirrhosis.[45]

Carbohydrate antigen 19-9 (CA19-9) and carcinoembryonic antigen (CEA) are the most widely used unspecific serum tumor markers, frequently found at high levels in patients with CCA, particularly at advance tumor stages.[47] However, they lack of sensitivity and specificity for the early diagnosis of CCA since their levels rarely overcome the standard cut-off value (37 IU/mL) at early tumor stages, thus questioning their usefulness as a general screening tool. Furthermore, high blood CA19-9 levels can also occur in other bile duct non-malignant diseases, such as bile duct inflammation and cholestasis.[48] Additionally, due to a deficiency in FUT3 activity, up to 7% of the general population are unable to express the CA19-9 epitope (*i.e.*, Lewis antigen negative individuals), limiting more its use as a diagnostic strategy.[48,49] Therefore, although the use of unspecific serum tumor markers for CCA diagnosis by themselves is not endorsed in routine clinical practice, CA19-9 and CEA results combined with imaging methods may support a suspected diagnosis of CCA.

Likewise, contrasting their diagnostic value, these serum tumor markers might be useful in estimating prognosis and/or treatment response. For instance, high pre-operative levels of CA19-9 (>200 IU/mL) may independently predict the prognosis of patients with CCA before tumor resection. Patients with resectable CCA and high CA19-9 levels may display a worse post-operative survival,[50] particularly if these levels do not normalize

after surgery.[51] Similarly, high CEA serum levels were shown to independently predict 1-, 3- and 5-year post-operative survival after tumor resection in patients with iCCA.[52,53] Regarding response to therapy, CA19-9 serum levels below 1000 IU/mL have been associated with a better response to gemcitabine-base therapeutic regimens and a reduction in serum CA19-9 levels in more than 50% during the treatment was reported to be a good indicator of prognosis.[54] Therefore, although these tumor biomarkers lack diagnostic accuracy at early stages, they may be used as prognostic biomarkers both at curative and palliative settings.

5.2 Imaging studies

Ordinarily, non-invasive abdominal imaging techniques to visualize the liver are routinely performed in order to try to find suspicious masses, followed by cholangiography imaging of the bile duct. Still, the diagnostic accuracy of imaging methods is influenced by tumor size, anatomical location and growth patterns of CCAs.[55]

5.2.1 Cross-sectional imaging

Ultrasonography (US), the technique that uses sound waves to view internal organs, often identifies bile duct dilatation and obstruction. According to the anatomical location of the tumor, US has generally a high detection rate for iCCA but there is also a risk of misclassification between iCCA and HCC.[56] In relation to extrahepatic bile duct malignancies, the accuracy of US for the identification of dCCA is around 80-95% while identifying pCCA is more challenging.[57] Endoscopic ultrasonography (EUS) and contrast-enhanced ultrasonography (CEUS) are UC-based modalities that can potentiate the diagnostic accuracy.

Computed tomography (CT) scanning is considered the standard imaging modality for the characterization, staging and resectability estimation of CCA by X-ray waves. Ductal dilatation and larger mass lesions are commonly detected by this method and tumor size can be non-invasively measured. However, when a lesion is smaller than 1 cm, or when facing liver cirrhosis or atypical features, the characterization is challenging.[57]

Magnetic resonance imaging (MRI) uses magnetic fields for the visualization and, compared to CT, it has a similar diagnostic accuracy, CCA staging capacity and a resectability estimation ability for pCCA.[58] To assess the presence of biliary tumors, comprehensive imaging protocols not only including the liver but also covering the biliary

tract and pancreas should be applied in order to rule out other malignancies such as pancreatic adenocarcinoma.[59] In this sense, for bile duct imaging, MRI allows doing magnetic resonance cholangiopancreatography (MRCP).

Positron emission tomography (PET), particularly ^{18}F-fluorodeoxyglucose PET (^{18}FDG-PET), is an accurate method for CCA tumor detection, for disease dissemination screening and for differentiation between benign and malignant structures. Even though, it may also result in false positive cases such as biliary inflammation or false-negative results, misdiagnosing mucinous tumors, for instance.[60] Consequently, PET is currently more used as the standard of care for disease staging and to identify tumor recurrence rather than for diagnostic purposes.[60]

5.2.2 Bile duct imaging (cholangiography)

Direct imaging of the bile ducts by cholangiograms is often necessary to confirm the presence of CCA. In routine clinical settings, endoscopic retrograde cholangiopancreatography (ERCP) and MRCP are the most used cholangiography techniques, though percuntaneous transhepatic cholangiography (PTC), is also becoming popular.[61]

ERCP allows the possibility to obtain tumor biopsies if the suspicious area is close to the small intestine as well as to place stents for relieving bile duct obstruction. Unlike ERCP, MRCP is a non-invasive imaging technique that allows the assessment of the biliary system with high accuracy.[62] PTC, which is the procedure based on the insertion of a needle into the bile duct, permits access to proximal lesions of both right and left ducts. Both ERCP and PTC can provide with material for cytologic/histologic studies.

5.3 Collection of histologic/cytologic samples for diagnostic testing

In order to get a histologic confirmation of CCA malignancy, biopsy/cytology is required. Biopsy/cytology samples are the only unequivocal way to confirm cancer presence in order to achieve a definitive diagnosis, for staging and also for detecting specific genetic alterations that may guide therapeutic decisions.[45,63] There are several ways to get a sample for testing but the type of sample-collecting methodology depends on tumor location. Additionally, biopsy sensitivity will depend on the tumor location, size, operator expertise and representativeness of the collected sample.[45]

ERCP with an inserted tiny brush enables to collect cells and tissue fragments for a biopsy. Via ERCP brush cytologies, biopsies, needle aspiration and/or shave biopsies may provide enough material for histologic studies.[64] PTC endobiliary forceps biopsy is also a valid procedure for histological assessment of proximal biliary strictures.[65] Although the use of PTC procedure is less common than ERCP, it offers a valuable alternative for patients with difficult bile duct access.

EUS-, CEUS- or CT-guided needle-aspiration also allows the acquisition of cytology samples for tumor testing. While ultrasonography with fine-needle aspiration has a greater sensitivity for detecting malignancy than ERCP with brush cytology, it increases the risk of tumor peritoneal seeding and patients with free-needle aspiration are illegible for liver transplantation.

The highest limitation of biopsy sampling procedures is sensibility, as the quality and quantity of cytological samples is usually quite poor, preventing negative cytological results from excluding CCA malignancy.[63,65] In this sense, fluorescence *in situ* hybridization (FISH) polysomy test in cytology samples has shown to increase the sensitivity for malignant stricture detection while maintaining high specificity. [66] FISH technique uses fluorescently labelled DNA probes to identify chromosomal aberrations in cells.

5.4 CCA diagnosis in patients with PSC

CCA screening among high-risk populations is currently conducted exclusively for patients with PSC. This screening is initially performed by measuring the levels of serum CA19-9 and by conducting MRCP or ultrasound, usually every 3-6 months. In this case, if biliary obstruction, elevated CA19-9 or a dominant stricture in MRCP results appear, subsequent ERCP with cytology/biopsy is recommended.[67–69]

The use of CA19-9 as a serum biomarker for CCA in PSC is limited as it is also frequently elevated due to cholangitis and cholestasis and is undetectable in Lewis-antigen-negative patients (around 7%).[49,70,71] Moreover, studies analysing CA19-9 levels in the context of CCA have reported different sensitivity and specificity values due to diverse cut-off levels among studies. A recent study using the international diagnostic cut-off of 37 U/mL provided a sensitivity of 17% and 93% of specificity for early CCA diagnosis in patients with PSC.[72] The low sensitivity of CA19-9 makes it a suboptimal marker for the early diagnosis of CCA.

Inflammatory biliary strictures occurring at the time of PSC presentation in 15-20% of cases mimic early malignant changes at imaging studies, such as ultrasound, CT and combined MRI/MRCP.

Radiological abnormalities observed using imaging techniques, including mass-forming lesions or thickening of the bile duct wall, might be indicative of CCA, but can also represent benign changes usually noticed in PSC, disabling their capacity to demonstrate the neoplastic nature of the stricture(s).[1,73,74]

Sensitivity and positive predictive value of brush cytology studies are rather poor for dominant strictures in PSC and advanced cytological techniques such as FISH polysomy detection lacks sensitivity and require invasive endoscopic procedures.[75,76]

Thus, the limited success of the tumor markers becomes a significant issue in PSC, whose clinical features and imaging findings overlap with those of CCA.

6. THERAPEUTIC STRATEGIES

6.1 Surgery

6.1.1 Surgical tumor resection

Complete tumor resection is the only therapy to afford a chance of cure for patients with CCA.[77] Unfortunately, due to suboptimal screening diagnostic tools and the lack of standardized screening programs for individuals at high risk, most patients (~90%) are not eligible for surgery at the time of diagnosis.[78,79] For this reason, the need to early diagnose CCA becomes imperative.

The main objective of tumor resection is to conduct a complete resection with negative macroscopic and microscopic margins (R0) and to ensure a viable liver remnant.[78]

Likewise, being CCA tumor relapse highly frequent after surgical resection, the use of adjuvant therapies after tumor resection to prevent tumor relapse has been established. In this sense, based on the benefits in terms of overall survival and relapse-free survival reported at the phase III BILCAP (NCT00363584) clinical trial, international guidelines recommend capecitabine as adjuvant therapy for 6 months after curative resection of CCA.[80]

6.1.2 Liver transplantation

The therapeutic benefit of orthotopic liver transplantation (LT) for intrahepatic or perihilar CCAs seems controversial based on past studies showing high recurrence-rates and low survival, as well as considering LT limitations associated to liver allograft supply and lifelong immunosuppression.[81,82]

However, recent studies have indicated that LT may offer an opportunity for early-stage but anatomically unresectable highly selected patients with pCCA.[67] Neoadjuvant chemoradiotherapy followed by LT resulted in a 5-year survival rate of 65-70%, with higher values among patients with PSC, who are often diagnosed earlier than *de novo* CCAs (5-year survival of 77%).[67] Hence, the success of LT as a therapeutic strategy for CCA may be attributed to early tumor diagnosis.

6.2 Systemic chemotherapy

The administration of systemic chemotherapy as a palliative treatment depends on patient's Eastern Cooperative Oncology Group Performance Status (ECOG-PS), disease distribution and accessibility of tumor profiling.[3]

The first-line chemotherapy treatment recommended by current international guidelines is the combination of gemcitabine and cisplatin (GemCis), as an increased overall survival was observed in the phase II ABC-02 (NCT00262769) and phase II BT22 clinical trials (NCT00380588).[83–85] If GemCis resistance develops overtime, FOLFOX (folinic acid, 5-fluorouracil and oxaliplatin) has presented some modest beneficial results in a phase III ABC-06 clinical trial (NCT01926236) as a second-line chemotherapeutic treatment.[86]

6.3 Locoregional therapies

Locoregional therapies such as transarterial chemoembolization (TACE), transarterial radioembolization with Yttrium^{90} (^{90}Y-TARE), selective internal radiation therapy (SIRT), radiofrequency ablation (RFA) and photodynamic therapy (PDT) have shown certain tumor reduction/control and increased survival, but the therapeutic value of these liver-directed therapies should be confirmed by future prospective studies.[87]

According to a previous meta-analysis including patients with unresectable CCA under palliative setting, the use of PDT as an adjunct to biliary stenting seemed to be more effective than biliary stenting alone, in terms of increased survival and decreased post-intervention cholangitis risk.[88]

Currently, the multicentre phase II ABC-02 study (EUDRACT 2014-003656-31) is recruiting patients with biliary tract cancer in order to study the potential additive benefit of stereotactic body radiotherapy (SBRT) to systemic chemotherapy in locally advanced cancers. This ongoing study will provide data on the potential benefits of this approach.

6.4 Targeted therapies

Attempting personalized medicine, several inhibitory molecules targeting some common genetic mutations in CCA have been developed and assayed. These targeted therapies are becoming available as second-line therapy for advanced CCAs with particular mutational signatures.

In this regard, Ivosidenib, an inhibitor of isocitrate dehydrogenase 1 (IDH1), has recently been approved by the FDA for treatment of adults with locally advanced or metastatic *IDH1*-mutated CCAs who progressed under first line GemCis treatment.[89,90]

For patients presenting fibroblast growth factor receptor *2 (FGFR2)* genetic fusions, two selective tyrosine kinase inhibitors have been recently approved. In 2020, the US FDA has approved the use of Pemigatinib for previously treated, unresectable, locally advanced or metastatic CCA with *FGFR2* fusions.[91] More recently, in 2021, the FDA has accelerated the approval of Infigratinib to be used in adults with previously treated, unresectable, locally advanced or metastatic CCA with *FGFR2* fusions.[92]

6.5 Immunotherapy

Fighting cancer through immune system activation seems to be encouraging as a potential therapeutic strategy in many cancers. So far, results from immune-directed therapies for CCAs are limited. Even so, some CAR T cell immunotherapy as well as immune checkpoint blockade with monoclonal antibodies (*i.e.*, anti-programmed cell death protein I (PD-1) antibodies) are being tested as an anti-cancer therapeutic option.[3]

Anti-tumor immune therapeutic strategies may lead to promising results in the future, but still warrant further investigation.

REFERENCES

1. Banales JM, Cardinale V, Carpino G, et al. Expert consensus document: Cholangiocarcinoma: current knowledge and future perspectives consensus statement from the European Network for the Study of Cholangiocarcinoma (ENS-CCA). *Nat Rev Gastroenterol Hepatol*. 2016;13(5):261-280. doi:10.1038/NRGASTRO.2016.51

2. Brindley PJ, Bachini M, Ilyas SI, et al. Cholangiocarcinoma. *Nat Rev Dis Prim*. 2021;7(1). doi:10.1038/S41572-021-00300-2

3. Banales JM, Marin JJG, Lamarca A, et al. Cholangiocarcinoma 2020: the next horizon in mechanisms and management. *Nat Rev Gastroenterol Hepatol*. 2020;17(9):557-588. doi:10.1038/S41575-020-0310-Z

4. Cardinale V, Carpino G, Reid L, Gaudio E. Multiple cells of origin in cholangiocarcinoma underlie biological, epidemiological and clinical heterogeneity. *World J Gastrointest Oncol*. 2012;4(5):94. doi:10.4251/wjgo.v4.i5.94

5. Fan B, Malato Y, Calvisi DF, et al. Cholangiocarcinomas can originate from hepatocytes in mice. *J Clin Invest*. 2012;122(8):2911-2915. doi:10.1172/JCI63212

6. Sekiya S, Suzuki A. Intrahepatic cholangiocarcinoma can arise from Notch-mediated conversion of hepatocytes. *J Clin Invest*. 2012;122(11):3914-3918. doi:10.1172/JCI63065

7. Fabris L, Perugorria MJ, Mertens J, et al. The tumour microenvironment and immune milieu of cholangiocarcinoma. *Liver Int*. 2019;39 Suppl 1(S1):63-78. doi:10.1111/LIV.14098

8. Blechacz B, Komuta M, Roskams T, Gores GJ. Clinical diagnosis and staging of cholangiocarcinoma. *Nat Rev Gastroenterol Hepatol*. 2011;8(9):512-522. doi:10.1038/NRGASTRO.2011.131

9. Klatskin G. Adenocarcinoma of the hepatic duct at its bifurcation within the porta hepatis. An unusual tumor with distinctive clinical and pathological features. *Am J Med*. 1965;38(2):241-256. doi:10.1016/0002-9343(65)90178-6

10. AJCC Cancer Staging Manual. *AJCC Cancer Staging Man*. Published online 2017. doi:10.1007/978-3-319-40618-3

11. Khan SA, Tavolari S, Brandi G. Cholangiocarcinoma: Epidemiology and risk factors. *Liver Int*. 2019;39 Suppl 1(S1):19-31. doi:10.1111/LIV.14095

12. Cardinale V, Semeraro R, Torrice A, et al. Intra-hepatic and extra-hepatic cholangiocarcinoma: New insight into epidemiology and risk factors. *World J Gastrointest Oncol*. 2010;2(11):407. doi:10.4251/WJGO.V2.I11.407

13. Khan SA, Emadossadaty S, Ladep NG, et al. Rising trends in cholangiocarcinoma: is the ICD classification system misleading us? *J Hepatol*. 2012;56(4):848-854. doi:10.1016/J.JHEP.2011.11.015

14. Bertuccio P, Malvezzi M, Carioli G, et al. Global trends in mortality from intrahepatic and extrahepatic cholangiocarcinoma. *J Hepatol*. 2019;71(1):104-114.

doi:10.1016/J.JHEP.2019.03.013

15. Baison GN, Bonds MM, Helton WS, Kozarek RA. Choledochal cysts: Similarities and differences between Asian and Western countries. *World J Gastroenterol*. 2019;25(26):3334-3343. doi:10.3748/WJG.V25.I26.3334

16. Clements O, Eliahoo J, Kim JU, Taylor-Robinson SD, Khan SA. Risk factors for intrahepatic and extrahepatic cholangiocarcinoma: A systematic review and meta-analysis. *J Hepatol*. 2020;72(1). doi:10.1016/J.JHEP.2019.09.007

17. Kim HJ, Kim JS, Joo MK, et al. Hepatolithiasis and intrahepatic cholangiocarcinoma: A review. *World J Gastroenterol*. 2015;21(48):13418-13431. doi:10.3748/WJG.V21.I48.13418

18. Karlsen TH, Folseraas T, Thorburn D, Vesterhus M. Primary sclerosing cholangitis - a comprehensive review. *J Hepatol*. 2017;67(6):1298-1323. doi:10.1016/J.JHEP.2017.07.022

19. Saffioti F, Mavroeidis VK. Review of incidence and outcomes of treatment of cholangiocarcinoma in patients with primary sclerosing cholangitis. *World J Gastrointest Oncol*. 2021;13(10):1336-1366. doi:10.4251/WJGO.V13.I10.1336

20. Boonstra K, Weersma RK, van Erpecum KJ, et al. Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. *Hepatology*. 2013;58(6):2045-2055. doi:10.1002/HEP.26565

21. Petrick JL, Yang B, Altekruse SF, et al. Risk factors for intrahepatic and extrahepatic cholangiocarcinoma in the United States: A population-based study in SEER-Medicare. *PLoS One*. 2017;12(10). doi:10.1371/JOURNAL.PONE.0186643

22. Aune D, Sen A, Norat T, Riboli E, Folseraas T. Primary sclerosing cholangitis and the risk of cancer, cardiovascular disease, and all-cause mortality: a systematic review and meta-analysis of cohort studies. *Sci Rep*. 2021;11(1). doi:10.1038/S41598-021-90175-W

23. Yonem O, Bayraktar Y. Clinical characteristics of Caroli's disease. *World J Gastroenterol*. 2007;13(13):1930-1933. doi:10.3748/WJG.V13.I13.1930

24. Kim TS, Pak JH, Kim JB, Bahk YY. Clonorchis sinensis, an oriental liver fluke, as a human biological agent of cholangiocarcinoma: a brief review. *BMB Rep*. 2016;49(11):590-597. doi:10.5483/BMBREP.2016.49.11.109

25. Sripa B, Kaewkes S, Sithithaworn P, et al. Liver fluke induces cholangiocarcinoma. *PLoS Med*. 2007;4(7):1148-1155. doi:10.1371/JOURNAL.PMED.0040201

26. Shin HR, Oh JK, Masuyer E, et al. Epidemiology of cholangiocarcinoma: an update focusing on risk factors. *Cancer Sci*. 2010;101(3):579-585. doi:10.1111/J.1349-7006.2009.01458.X

27. Ralphs S, Khan SA. The role of the hepatitis viruses in cholangiocarcinoma. *J Viral Hepat*. 2013;20(5):297-305. doi:10.1111/JVH.12093

28. Allen KJ, Gurrin LC, Constantine CC, et al. Iron-Overload–Related Disease in HFE Hereditary Hemochromatosis . *N Engl J Med*. 2008;358(3):221-230. doi:10.1056/nejmoa073286

29. Beyer G, Habtezion A, Werner J, Lerch MM, Mayerle J. Chronic pancreatitis. *Lancet (London, England)*. 2020;396(10249):499-512. doi:10.1016/S0140-6736(20)31318-0

30. Jing W, Jin G, Zhou X, et al. Diabetes mellitus and increased risk of cholangiocarcinoma:

a meta-analysis. *Eur J Cancer Prev*. 2012;21(1):24-31. doi:10.1097/CEJ.0B013E3283481D89

31. Byrne CD, Targher G. NAFLD: a multisystem disease. *J Hepatol*. 2015;62(1 Suppl):S47-S64. doi:10.1016/J.JHEP.2014.12.012

32. Wongjarupong N, Assavapongpaiboon B, Susantitaphong P, et al. Non-alcoholic fatty liver disease as a risk factor for cholangiocarcinoma: a systematic review and meta-analysis. *BMC Gastroenterol*. 2017;17(1):149. doi:10.1186/S12876-017-0696-4

33. Osataphan S, Mahankasuwan T, Saengboonmee C. Obesity and cholangiocarcinoma: A review of epidemiological and molecular associations. *J Hepatobiliary Pancreat Sci*. Published online 2021. doi:10.1002/JHBP.1001

34. Palmer WC, Patel T. Are common factors involved in the pathogenesis of primary liver cancers? A meta-analysis of risk factors for intrahepatic cholangiocarcinoma. *J Hepatol*. 2012;57(1):69-76. doi:10.1016/J.JHEP.2012.02.022

35. Petrick JL, Thistle JE, Zeleniuch-Jacquotte A, et al. Body Mass Index, Diabetes and Intrahepatic Cholangiocarcinoma Risk: The Liver Cancer Pooling Project and Meta-analysis. *Am J Gastroenterol*. 2018;113(10):1494-1505. doi:10.1038/S41395-018-0207-4

36. Li JS, Han TJ, Jing N, et al. Obesity and the risk of cholangiocarcinoma: a meta-analysis. *Tumour Biol*. 2014;35(7):6831-6838. doi:10.1007/S13277-014-1939-4

37. Huang Y, You L, Xie W, Ning L, Lang J. Smoking and risk of cholangiocarcinoma: a systematic review and meta-analysis. *Oncotarget*. 2017;8(59):100570-100581. doi:10.18632/ONCOTARGET.20141

38. Travis LB, Hauptmann M, Gaul LK, et al. Site-specific cancer incidence and mortality after cerebral angiography with radioactive thorotrast. *Radiat Res*. 2003;160(6):691-706. doi:10.1667/RR3095

39. Ishikawa Y, Wada I, Fukumoto M. Alpha-particle carcinogenesis in Thorotrast patients: epidemiology, dosimetry, pathology, and molecular analysis. *J Environ Pathol Toxicol Oncol Off organ Int Soc Environ Toxicol Cancer*. 2001;20(4):311-315.

40. Kato I, Kido C. Increased risk of death in thorotrast-exposed patients during the late follow-up period. *Jpn J Cancer Res*. 1987;78(11):1187-1192.

41. Kumagai S, Sobue T, Makiuchi T, et al. Relationship between cumulative exposure to 1,2-dichloropropane and incidence risk of cholangiocarcinoma among offset printing workers. *Occup Environ Med*. 2016;73(8):545-552. doi:10.1136/oemed-2015-103427

42. Brandi G, Girolamo S Di, Farioli A, et al. Asbestos: a hidden player behind the cholangiocarcinoma increase? Findings from a case-control analysis. *Cancer Causes Control*. 2013;24(5):911-918. doi:10.1007/S10552-013-0167-3

43. Farioli A, Straif K, Brandi G, et al. Occupational exposure to asbestos and risk of cholangiocarcinoma: a population-based case-control study in four Nordic countries. *Occup Environ Med*. 2018;75(3):191-198. doi:10.1136/OEMED-2017-104603

44. Rodrigues PM, Olaizola P, Paiva NA, et al. Pathogenesis of Cholangiocarcinoma. *Annu Rev Pathol*. 2021;16:433-463. doi:10.1146/ANNUREV-PATHOL-030220-020455

45. Forner A, Vidili G, Rengo M, Bujanda L, Ponz-Sarvisé M, Lamarca A. Clinical presentation, diagnosis and staging of cholangiocarcinoma. *Liver Int*. 2019;39 Suppl 1(S1):98-107. doi:10.1111/LIV.14086

46. Cardinale V, Bragazzi MC, Carpino G, et al. Intrahepatic cholangiocarcinoma: review and update. *Hepatoma Res*. 2018;4:20. doi:10.20517/2394-5079.2018.46

47. Macias RIR, Banales JM, Sangro B, et al. The search for novel diagnostic and prognostic biomarkers in cholangiocarcinoma. *Biochim Biophys acta Mol basis Dis*. 2018;1864(4 Pt B):1468-1477. doi:10.1016/J.BBADIS.2017.08.002

48. Onal C, Colakoglu T, Ulusan SN, Yapar AF, Kayaselcuk F. Biliary obstruction induces extremely elevated serum CA 19-9 levels: case report. *Onkologie*. 2012;35(12):780-782. doi:10.1159/000345110

49. Wannhoff A, Hov JR, Folseraas T, et al. FUT2 and FUT3 genotype determines CA19-9 cut-off values for detection of cholangiocarcinoma in patients with primary sclerosing cholangitis. *J Hepatol*. 2013;59(6):1278-1284. doi:10.1016/J.JHEP.2013.08.005

50. Kondo N, Murakami Y, Uemura K, et al. Elevated perioperative serum CA 19-9 levels are independent predictors of poor survival in patients with resectable cholangiocarcinoma. *J Surg Oncol*. 2014;110(4):422-429. doi:10.1002/JSO.23666

51. Yamashita S, Passot G, Aloia TA, et al. Prognostic value of carbohydrate antigen 19-9 in patients undergoing resection of biliary tract cancer. *Br J Surg*. 2017;104(3):267-277. doi:10.1002/BJS.10415

52. Luo X, Yuan L, Wang Y, Ge R, Sun Y, Wei G. Survival outcomes and prognostic factors of surgical therapy for all potentially resectable intrahepatic cholangiocarcinoma: a large single-center cohort study. *J Gastrointest Surg*. 2014;18(3):562-572. doi:10.1007/S11605-013-2447-3

53. Qiang Z, Zhang W, Jin S, et al. Carcinoembryonic antigen, α-fetoprotein, and Ki67 as biomarkers and prognostic factors in intrahepatic cholangiocarcinoma: A retrospective cohort study. *Ann Hepatol*. 2021;20. doi:10.1016/J.AOHEP.2020.07.010

54. Lee BS, Lee SH, Son JH, et al. Prognostic value of CA 19-9 kinetics during gemcitabine-based chemotherapy in patients with advanced cholangiocarcinoma. *J Gastroenterol Hepatol*. 2016;31(2):493-500. doi:10.1111/JGH.13059

55. Joo I, Lee JM, Yoon JH. Imaging Diagnosis of Intrahepatic and Perihilar Cholangiocarcinoma: Recent Advances and Challenges. *Radiology*. 2018;288(1):7-23. doi:10.1148/RADIOL.2018171187

56. Vilana R, Forner A, Bianchi L, et al. Intrahepatic peripheral cholangiocarcinoma in cirrhosis patients may display a vascular pattern similar to hepatocellular carcinoma on contrast-enhanced ultrasound. *Hepatology*. 2010;51(6):2020-2029. doi:10.1002/HEP.23600

57. Mar WA, Shon AM, Lu Y, et al. Imaging spectrum of cholangiocarcinoma: role in diagnosis, staging, and posttreatment evaluation. *Abdom Radiol (New York)*. 2016;41(3):553-567. doi:10.1007/S00261-015-0583-9

58. Lo EC, N. Rucker A, Federle MP. Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma: Imaging for Diagnosis, Tumor Response to Treatment and Liver Response to Radiation. *Semin Radiat Oncol*. 2018;28(4):267-276. doi:10.1016/J.SEMRADONC.2018.06.010

59. Vanderveen KA, Hussain HK. Magnetic Resonance Imaging of cholangiocarcinoma. *Cancer Imaging*. 2004;4(2):104-115. doi:10.1102/1470-7330.2004.0018

60. Lamarca A, Barriuso J, Chander A, et al. 18 F-fluorodeoxyglucose positron emission tomography (18 FDG-PET) for patients with biliary tract cancer: Systematic review and

meta-analysis. *J Hepatol*. 2019;71(1):115-129. doi:10.1016/J.JHEP.2019.01.038

61. Florescu LM, Florescu DN, Gheonea IA. The Importance of Imaging Techniques in the Assessment of Biliary Tract Cancer. *Curr Heal Sci J*. 2017;43(3):201-208. doi:10.12865/CHSJ.43.03.03

62. Miletić D, Štimac D, Uravić M, et al. Magnetic resonance cholangiopancreatography: a meta-analysis of test performance in suspected biliary disease. *Ann Intern Med*. 2003;139(7):336-343. doi:10.7326/0003-4819-139-7-200310070-00006

63. Bridgewater J, Galle PR, Khan SA, et al. Guidelines for the diagnosis and management of intrahepatic cholangiocarcinoma. *J Hepatol*. 2014;60(6):1268-1289. doi:10.1016/J.JHEP.2014.01.021

64. Coelho-Prabhu N, Baron TH. Endoscopic retrograde cholangiopancreatography in the diagnosis and management of cholangiocarcinoma. *Clin Liver Dis*. 2010;14(2):333-348. doi:10.1016/J.CLD.2010.03.011

65. Weber A, Schmid RM, Prinz C. Diagnostic approaches for cholangiocarcinoma. *World J Gastroenterol*. 2008;14(26):4131-4136. doi:10.3748/WJG.14.4131

66. Liew ZH, Loh TJ, Lim TKH, et al. Role of fluorescence in situ hybridization in diagnosing cholangiocarcinoma in indeterminate biliary strictures. *J Gastroenterol Hepatol*. 2018;33(1):315-319. doi:10.1111/JGH.13824

67. Zamora-Valdes D, Heimbach JK. Liver Transplant for Cholangiocarcinoma. *Gastroenterol Clin North Am*. 2018;47(2):267-280. doi:10.1016/J.GTC.2018.01.002

68. Taghavi SA, Eshraghian A, Niknam R, Sivandzadeh GR, Bagheri Lankarani K. Diagnosis of cholangiocarcinoma in primary sclerosing cholangitis. *Expert Rev Gastroenterol Hepatol*. 2018;12(6):575-584. doi:10.1080/17474124.2018.1473761

69. Vedeld HM, Folseraas T, Lind GE. Detecting cholangiocarcinoma in patients with primary sclerosing cholangitis - The promise of DNA methylation and molecular biomarkers. *JHEP reports Innov Hepatol*. 2020;2(5). doi:10.1016/J.JHEPR.2020.100143

70. Grimsrud MM, Folseraas T. Pathogenesis, diagnosis and treatment of premalignant and malignant stages of cholangiocarcinoma in primary sclerosing cholangitis. *Liver Int*. 2019;39(12):2230-2237. doi:10.1111/LIV.14180

71. Fisher A, Theise ND, Min A, et al. CA19-9 does not predict cholangiocarcinoma in patients with primary sclerosing cholangitis undergoing liver transplantation. *Liver Transpl Surg*. 1995;1(2):94-98. doi:10.1002/LT.500010204

72. Cuenco J, Wehnert N, Blyuss O, et al. Identification of a serum biomarker panel for the differential diagnosis of cholangiocarcinoma and primary sclerosing cholangitis. *Oncotarget*. 2018;9(25):17430-17442. doi:10.18632/ONCOTARGET.24732

73. Rizvi S, Eaton JE, Gores GJ. Primary Sclerosing Cholangitis as a Premalignant Biliary Tract Disease: Surveillance and Management. *Clin Gastroenterol Hepatol*. 2015;13(12):2152-2165. doi:10.1016/J.CGH.2015.05.035

74. Njei B, McCarty TR, Varadarajulu S, Navaneethan U. Systematic review with meta-analysis: endoscopic retrograde cholangiopancreatography-based modalities for the diagnosis of cholangiocarcinoma in primary sclerosing cholangitis. *Aliment Pharmacol Ther*. 2016;44(11-12):1139-1151. doi:10.1111/APT.13817

75. Ponsioen CY, Vrouenraets SME, Van Milligen De Wit AWM, et al. Value of brush cytology

for dominant strictures in primary sclerosing cholangitis. *Endoscopy*. 1999;31(4):305-309. doi:10.1055/S-1999-18

76. Navaneethan U, Njei B, Venkatesh PGK, Vargo JJ, Parsi MA. Fluorescence in situ hybridization for diagnosis of cholangiocarcinoma in primary sclerosing cholangitis: a systematic review and meta-analysis. *Gastrointest Endosc*. 2014;79(6). doi:10.1016/J.GIE.2013.11.001

77. Nathan H, Pawlik TM, Wolfgang CL, Choti MA, Cameron JL, Schulick RD. Trends in survival after surgery for cholangiocarcinoma: a 30-year population-based SEER database analysis. *J Gastrointest Surg*. 2007;11(11):1488-1497. doi:10.1007/S11605-007-0282-0

78. Radtke A, Königsrainer A. Surgical Therapy of Cholangiocarcinoma. *Visc Med*. 2016;32(6):422-426. doi:10.1159/000452921

79. Valle JW, Borbath I, Khan SA, et al. Biliary cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol Off J Eur Soc Med Oncol*. 2016;27(suppl 5):v28-v37. doi:10.1093/ANNONC/MDW324

80. Shroff RT, Kennedy EB, Bachini M, et al. Adjuvant Therapy for Resected Biliary Tract Cancer: ASCO Clinical Practice Guideline. *J Clin Oncol*. 2019;37(12):1015-1027. doi:10.1200/JCO.18.02178

81. Rosen CB, Heimbach JK, Gores GJ. Liver transplantation for cholangiocarcinoma. *Transpl Int*. 2010;23(7):692-697. doi:10.1111/J.1432-2277.2010.01108.X

82. Gringeri E, Gambato M, Sapisochin G, et al. Cholangiocarcinoma as an Indication for Liver Transplantation in the Era of Transplant Oncology. *J Clin Med*. 2020;9(5). doi:10.3390/JCM9051353

83. Valle J, Wasan H, Palmer DH, et al. Cisplatin plus gemcitabine versus gemcitabine for biliary tract cancer. *N Engl J Med*. 2010;362(14):1273-1281. doi:10.1056/NEJMOA0908721

84. Okusaka T, Nakachi K, Fukutomi A, et al. Gemcitabine alone or in combination with cisplatin in patients with biliary tract cancer: a comparative multicentre study in Japan. *Br J Cancer*. 2010;103(4):469-474. doi:10.1038/SJ.BJC.6605779

85. Thongprasert S, Napapan S, Charoentum C, Moonprakan S. Phase II study of gemcitabine and cisplatin as first-line chemotherapy in inoperable biliary tract carcinoma. *Ann Oncol Off J Eur Soc Med Oncol*. 2005;16(2):279-281. doi:10.1093/ANNONC/MDI046

86. Lamarca A, Palmer DH, Wasan HS, et al. Second-line FOLFOX chemotherapy versus active symptom control for advanced biliary tract cancer (ABC-06): a phase 3, open-label, randomised, controlled trial. *Lancet Oncol*. 2021;22(5):690-701. doi:10.1016/S1470-2045(21)00027-9

87. Renzulli M, Ramai D, Singh J, et al. Locoregional Treatments in Cholangiocarcinoma and Combined Hepatocellular Cholangiocarcinoma. *Cancers (Basel)*. 2021;13(13). doi:10.3390/CANCERS13133336

88. Moole H, Tathireddy H, Dharmapuri S, et al. Success of photodynamic therapy in palliating patients with nonresectable cholangiocarcinoma: A systematic review and meta-analysis. *World J Gastroenterol*. 2017;23(7):1278-1288. doi:10.3748/WJG.V23.I7.1278

89. Ivosidenib Boosts OS in Cholangiocarcinoma. *Cancer Discov*. Published online October 11, 2021. doi:10.1158/2159-8290.CD-NB2021-0389

90. Zhu AX, Macarulla T, Javle MM, et al. Final Overall Survival Efficacy Results of Ivosidenib for Patients With Advanced Cholangiocarcinoma With IDH1 Mutation: The Phase 3 Randomized Clinical ClarIDHy Trial. *JAMA Oncol*. 2021;7(11). doi:10.1001/JAMAONCOL.2021.3836

91. Abou-Alfa GK, Sahai V, Hollebecque A, et al. Pemigatinib for previously treated, locally advanced or metastatic cholangiocarcinoma: a multicentre, open-label, phase 2 study. *Lancet Oncol*. 2020;21(5):671-684. doi:10.1016/S1470-2045(20)30109-1

92. Javle M, Roychowdhury S, Kelley RK, et al. Infigratinib (BGJ398) in previously treated patients with advanced or metastatic cholangiocarcinoma with FGFR2 fusions or rearrangements: mature results from a multicentre, open-label, single-arm, phase 2 study. *lancet Gastroenterol Hepatol*. 2021;6(10):803-815. doi:10.1016/S2468-1253(21)00196-5

CHAPTER 2

EXTRACELLULAR VESICLES IN HEPATOBILIARY MALIGNANCIES

Ainhoa Lapitz,[1] Ander Arbelaiz,[1] Paula Olaizola,[1] Aitziber Aranburu,[1] Luis Bujanda,[1,2] Maria J Perugorria,[1,2,3] Jesus M Banales[1,2,3]

[1]Department of Liver and Gastrointestinal Diseases, Biodonostia Research Institute, Donostia University Hospital, University of the Basque Country (UPV/EHU), San Sebastian, Spain.

[2]"Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas" (CIBERehd), Carlos III National Institute of Health, Madrid, Spain.

[3]IKERBASQUE, Basque Foundation for Science, Bilbao, Spain.

Published in *Frontiers in Immunology, 2018 Oct 12;9:2270*

ABSTRACT

Primary hepatobiliary malignancies include a heterogeneous group of cancers with dismal prognosis, among which hepatocellular carcinoma (HCC), cholangiocarcinoma (CCA), and hepatoblastoma (HB) stand out. These tumors mainly arise from the malignant transformation of hepatocytes, cholangiocytes (bile duct epithelial cells) or hepatoblasts (embryonic liver progenitor cells), respectively. Early diagnosis, prognosis prediction and effective therapies are still a utopia for these diseases. Extracellular vesicles (EVs) are small membrane-enclosed spheres secreted by cells and present in biological fluids. They contain multiple types of biomolecules, such as proteins, RNA, DNA, metabolites and lipids, which make them a potential source of biomarkers as well as regulators of human pathobiology. In this review, the role of EVs in the pathogenesis of hepatobiliary cancers and their potential usefulness as disease biomarkers are highlighted. Moreover, the therapeutic value of EV regulation is discussed and future directions on basic and clinical research are indicated.

1. INTRODUCTION

Liver cancer is a major health problem worldwide, representing the second leading cause of all cancer-related deaths.[1] This cancer involves a heterogeneous set of hepatobiliary malignancies including hepatocellular carcinoma (HCC), cholangiocarcinoma (CCA), and hepatoblastoma (HB), which mainly arise from the malignant transformation of hepatocytes, cholangiocytes, and hepatoblasts, respectively.[1,2] Early non-invasive diagnosis, prediction of prognosis and treatment-response, as well as effective personalized therapies are still a challenge, highly compromising patient outcome.[3–5]

HCC is the sixth most prevalent malignant tumor (10:100,000 incidence) and is strongly associated (~90%) with the presence of liver cirrhosis (LC) caused by alcohol, viral infections [hepatitis B (HBV) or C (HCV) viruses], and/or steatosis, among others.[5,6] CCA is a rare cancer, but its incidence (~5/100.000) is increasing worldwide. Although the etiology of the majority of CCAs is unknown, several risk factors may predispose for its development, including the presence of primary sclerosing cholangitis (PSC), liver fluke infections (endemic from East Asia), cirrhosis and congenital biliary disorders.[3] On the other hand, HB is the most common pediatric liver malignancy, principally affecting children between 6 months and 3 years of age. HB is responsible for up to ~1% of all pediatric cancers, with an annual incidence of 0.5–1.5 cases.[4,7] Despite most HB cases are sporadic, some of them have been associated with hereditary cancer syndromes including familial adenomatous polyposis (FAP) and Beckwith-Widemann syndrome (BWS), as well as with prematurity or low birth weight.[4,7] Since hepatobiliary malignancies are usually diagnosed in late stages and are highly chemoresistant, the complete surgical resection of the tumors or liver transplantation constitute the only potential curative options. However, these therapeutic strategies are exclusively indicated under certain strict and conservative clinical criteria.[3,5,6] Therefore, there is an urgent need to determine new accurate non-invasive biomarkers for the early diagnosis of these diseases, as well as to monitor and predict disease progression and treatment response. Moreover, new effective personalized treatments are desirable in order to improve the outcome and life quality of patients.

During the last decade, extracellular vesicles (EVs) have opened new opportunities for non-invasive diagnosis and monitoring of human diseases. Their presence in biological fluids (serum, urine, bile, saliva, etc.) and their unique and diverse biomolecular composition (proteins, RNA, DNA, metabolites, and lipids) make EVs excellent candidates as a source of biomarkers.[8,9] Furthermore, since EVs participate in intercellular communication in human health and disease, they have been postulated as

potential tools or targets for therapy. EVs are small membrane-encapsulated spheres produced and secreted by cells through complex and precise molecular mechanisms.[10–13] Traditionally, EVs are classified according to their biogenesis into exosomes, microvesicles (MVs) or microparticles, and apoptotic bodies.[11,12] Exosomes are referred to those EVs produced inside the multivesicular endosomes (MVEs) of the cells. Their morphology is spherical and the size ranges between 40 and 150 nm in diameter.[11,14,15] Cell MVEs are vesicular entities generated in the maturation process of the early endosomes, and where intraluminal vesicles (ILVs) are formed by the invagination of the MVE membrane. ILVs are the incipient exosomes that are released to the extracellular media upon the fusion of the MVEs with the plasma membrane of the cell.[11] On the other hand, MVs or microparticles originate from the direct budding of the cell plasma membrane. Their size (40–1000 nm) and morphology are heterogeneous.[15,16] Apoptotic bodies are vesicles produced by cells undergoing apoptosis. Thus, their size (~40 to 2000–5000 nm) and morphology are diverse.[15,16] Although this classification is widely accepted, to date, there are no specific biomarkers to differentiate exosomes from other types of nano-sized vesicle populations, limiting their specific isolation from biofluids. Therefore, the smallest vesicles (nano-vesicles) present and isolated from biological fluids comprise a mix of exosomes and plasma membrane-derived vesicles.

EVs have changed the paradigm of intercellular communication, which was traditionally restricted to the autocrine, paracrine and endocrine interaction through soluble proteins and lipids, or through direct cell-to-cell contact mediated by proteins, gap junctions, or tunnelling nano-tubes in pluricellular organisms.[17,18] Accordingly, EVs contain an aqueous lumen and a specific subset of membrane and soluble proteins, nucleic acids (DNA and RNA), lipids and metabolites that can be horizontally transferred to local or distant cells by direct EV-cell membrane contact, fusion or internalization.[12] The importance of EVs is highlighted by the fact that their composition is specific depending on the cell status and on the received stimuli,[19,20] which indicates a certain degree of selective packaging. EVs confer protection to the biomolecules enclosed inside the lipid bilayer, preventing their enzymatic degradation.[21]

EVs participate in the regulation of multiple cancer hallmarks. They can transmit oncogenic signals by transferring pro-tumor RNAs and proteins that regulate diverse key processes in tumor progression such as proliferation, survival, differentiation, and invasion/migration of cancer cells.[22–30] Additionally, EVs are involved in the crosstalk between tumor cells and stroma, promoting inflammation,[31] cell matrix remodeling,[32] neovascularization or angiogenesis,[33,34] chemoresistance,[35–37] formation of the metastatic niche,[38,39] and inhibition of the anti-tumor immune response.[40–42] Therefore,

EVs represent key targets for therapy at various levels, including production, release and uptake by target cells. Additionally, the blockage or removal of tumor-derived EVs by apheresis with specific devices constitutes a potential therapeutic approach. Of note, EVs are also excellent candidates for the delivery of new anti-cancer proteins, RNAs, metabolites, drugs or cancer vaccines.

2. EV BIOGENESIS AND REGULATION

EV production is a highly regulated and complex cellular process where several protein networks and diverse intracellular signals are involved (**Figure 2.1**). Among the different EV populations, the exosome production machinery is the best studied. Nevertheless, both exosomes and MVs share numerous mechanisms that participate in their biogenesis, release, and uptake.[43] Regarding the mechanisms involved in the biogenesis of exosomes, the endosomal sorting complex required for transport (ESCRT) machinery has been reported to participate in MVE and ILV generation.[44] However, ESCRT independent mechanisms have also been described in the formation of exosomes, including ceramide production by neutral type II sphingomyelinase (nSMase2),[45,46] lipid rafts,[47] phospholipase D2 (PLD2),[48,49] and tetraspanin family of proteins (*e.g.*, CD9, CD63, and CD81),[50,51] which form dynamic membrane microdomains that promote their budding and assure exosome formation.

Intracellular trafficking of MVEs is coordinated by the cytoskeleton and motor proteins such as dynein and Rab family of GTPases (**Figure 2.1**).[52,53] MVEs fuse with the plasma membrane via SNARE proteins, finally allowing exosome secretion.[53,54] In contrast, MVs arise as a result of the direct budding of the cell plasma membrane. Their biogenesis requires Ca^{2+}-dependent membrane phospholipid and cytoskeleton rearrangements, which enable MV blebbing and release.[55] Released EVs are recognized by the recipient cells through specific interactions between their membrane components. These include integrins, lipids, tetraspanins, proteoglycans, among others.[56] Plasma membrane-bound EVs can be internalized through clathrin-mediated endocytosis (CME) or clathrin-independent processes that include phagocytosis, macropinocytosis and lipid rafts.

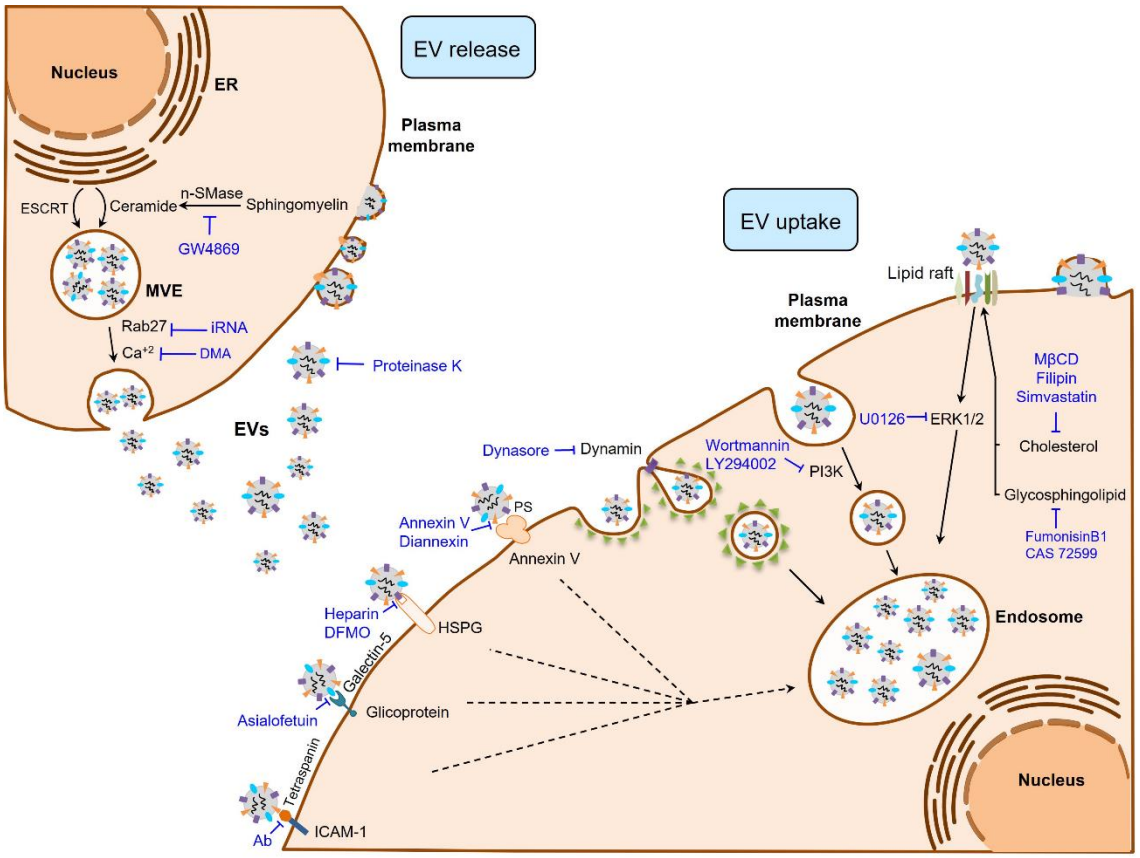


Figure 2.1: **Regulatory mechanisms of EV biogenesis, release, and uptake.** Exosome release can be inhibited by interfering their biogenesis (*e.g.*, ceramide production) or the membrane fusion of the multivesicular endosome (MVE) with the plasma membrane (*e.g.*, Rab27). Once EVs are released to the extracellular milieu, their uptake can be blocked by interfering the EV-plasma membrane protein interactions (*e.g.*, Tetraspanins), clathrin- and caveolin-dependent endocytosis (*e.g.*, Dynasore), phagocytosis (*e.g.*, Wortmannin), and by inhibiting lipid-raft mediated endocytosis (*e.g.*, Filipin). DFMO, difluoromethylornithine; DMA, dimethyl amiloride; ESCRT, endosomal sorting complex required for transport; EVs, extracellular vesicles; HSPG, heparan sulfate proteoglycans; ICAM-1, intercellular adhesion molecule 1; MβCD, methyl-β-cyclodextrin; nSMase, neutral sphingomyelinase; PS, phosphatidylserine.

Several experimental strategies have targeted the aforementioned mechanisms to interfere with EV production, release and uptake at different levels (**Figure 2.1**). Among these, intervention on ceramide production has been the most widely used strategy to decrease exosome production in cancer cells and thereby abolish the multiple oncogenic effects of tumor-derived exosomes in several cancers. Experimental inhibition of nSMase2, responsible for ceramide production, with its inhibitor GW4869 reduces exosome secretion and sensitizes cancer cells to chemotherapy.[45,57,58] Of note, the presence of GW4869 inhibits the migratory capacity of CCA cells.[31] Several intercellular signals involved in the regulation of the EV production are also under investigation, including the reduction of intracellular Ca^{2+} concentration. In fact, the Na^{+}/Ca^{+2} exchange inhibitor dimethyl amiloride (DMA) leads to diminished EV production in lymphoma cells,

resulting in an enhanced anti-tumor immune response.[40] Regarding the proteins that participate in the transport of exosomes, Rab family proteins are key mediators of MVE transit to the plasma membrane, their inhibition being linked to a decrease of exosome release.[59] Accordingly, the repression of Rab27a diminished growth and dissemination of cancer cells in vivo.[39,60]

Different complex mechanisms, including protein and/or lipid interactions between EV and recipient cell surface components, are required for the EV cell uptake (**Figure 2.1**). These mechanisms include phagocytosis, macropinocytosis, clathrin, and caveolin dependent endocytic pathways, as well as lipid raft-mediated and membrane fusion processes.[56] Therefore, aiming to block tumor EV uptake processes, different components of these machineries have been targeted. Membrane proteins including tetraspanins, integrins, lectins, proteoglycans, major histocompatibility complex (MHC) molecules, glycoproteins and other receptors are involved in EV-recipient cell interaction. Tetraspanins, enriched proteins present in EVs and well-established markers of these vesicles,[61,62] participate in EV-cell surface adhesion mediating their uptake. Thus, antibody-based inhibition of the CD81 and CD9 tetraspanins as well as the blockade of αV and β3 integrins, hampers EV uptake.[63,64] Lectins, such as galectin-5, can also be targeted with the glycoprotein asialofetuin to interfere with EV-cell interaction and the subsequent cellular internalization.[65] Likewise, targeting the lectin receptors DC-SIGN or DEC-205 with specific antibodies also results in a reduction of EV uptake.[66,67] Heparin can also block the internalization of cancer EVs by binding to the cell surface heparan sulfate proteoglycans.[68,69] Furthermore, the pivotal interaction between EVs and the plasma membrane of the cell can be partially inhibited by proteinase K treatment, blocking EV recognition and the subsequent endocytic process in cancer cells.[70]

The best studied EV internalization mechanisms are related with the endocytic pathway.[63,70,71] Since these processes depend on the cytoskeleton, the inhibition of actin polymerization by cytochalasin D reduces EV uptake by phagocytosis.[56,70,71] In addition, the uptake of EVs by macrophages can be abrogated by the inhibition of phosphoinositide-3-kinase (PI3K) with wortmannin or LY294002.[72] Moreover, inhibitors of macropinocytosis or CME [5-ethyl-N-isopropyl amiloride (EIPA) and chlorpromazine, respectively] reduce tumor-derived EV internalization.[70] Dynasore can also impair CME through the inhibition of dynamin 2, needed for clathrin-coated endosome membrane fission.[65,73–76] On the other hand, certain endocytic processes are closely related to lipid rafts, and the intervention on their composition impairs the uptake of EVs. Thus, the use of the glycosphingolipid synthesis inhibitors [*i.e.*, fumosinin B1 and N-butyldeoxynojirimycin hydrochloride (also known as CAS72599)] reduces EV uptake.[77]

Cholesterol reducing agents including methyl-beta-cyclodextrin (MβCD),[70,78,79] filipin [71,79] and simvastatin, as well as the inhibition of ERK1/2 signalling by U0126 may also impair the uptake of EVs.[79] Masking phosphatidylserine (PS), present in the membrane surface of EVs, by Diannexin and blocking its receptor TIM4 inhibits epidermal growth factor receptor (EGFR) transfer from tumor EVs to endothelial cells resulting in reduced tumor growth and microvascular density in vivo.[80] EVs can also release their content into the recipient cells by the direct fusion of plasma and EV membranes. This fusion is enhanced in acidic conditions, a general feature of cancer cells.[81] In this sense, proton pump inhibition leads to reduced EV uptake by cancer cells.[82]

3. EVS IN HEPATOBILIARY CANCERS

3.1 Hepatocellular Carcinoma

HCC cell-derived EVs participate in autocrine and/or paracrine cellular communications, regulating tumor growth, chemoresistance, angiogenesis, and dissemination. Several lines of evidence indicate that HCC cell-derived EVs promote tumor resistance against chemotherapeutic drugs such as sorafenib, doxorubicin or camptothecin. For instance, an enrichment of long intergenic non-coding RNA regulator of reprogramming (linc-ROR) in EVs derived from sorafenib-treated HCC cells prevents chemotherapy-induced apoptosis through p53 repression and increases the expression of tumor-initiating liver cancer stem cell CD133 marker.[83] Another molecular mechanism involved in HCC cell-derived EV-induced sorafenib resistance includes the activation of the hepatocyte growth factor (HGF)/c-Met/Akt signalling pathway in liver cancer cells.[84]

EVs derived from HCC cells may also regulate angiogenesis.[85] Experimental in vitro models indicate that EVs derived from CD90+ liver cancer cells (*i.e.*, cancer stem-like cells present in primary tumors and blood of HCC patients, associated with metastasis as well as bad prognosis) are enriched in long non-coding RNA (lncRNA) H19, which promotes the expression of vascular endothelial growth factor (VEGF) and its receptor VEGF-R1 in endothelial cells. Moreover, lncRNA H19 stimulates tube formation as well as cell-adhesion properties in endothelial cells, inducing the expression of intercellular adhesion molecule 1 (ICAM-1) in this cell-type.

Concerning the role of EVs in metastasis, several studies have reported that EVs secreted from HCC cells or from adjacent cells are also involved in the promotion of tumor cell metastasis.[86] Transcriptomic and proteomic profiling revealed that EVs derived

from metastatic HCC cells carry a larger number of pro-tumorigenic RNAs and proteins, such as MET proto-oncogene, S100 family members (S100A4, S100A10, and S100A11) and the caveolins (CAV1 and CAV2). HCC-derived EVs trigger the activation of PI3K/Akt and MAPK signalling pathways and the secretion of active MMP2 and MMP9 matrix metalloproteinases (MMPs) in hepatocytes, which in turn enhance their migratory and invasive ability. On the other hand, cancer-associated fibroblast (CAF)-derived EVs may also contribute to HCC cell proliferation and metastasis.[87] Thus, a reduction in the miR-320a level was observed in CAF-derived EVs compared to para-cancer fibroblasts (PAFs). This miR-320a directly targets pre-B-cell leukemia transcription factor 3 (PBX3), suppressing HCC cell proliferation, migration, and invasion. The anti-tumor effects of miR-320a were confirmed in vivo using HCC tumor xenograft models, in which tumor growth was inhibited when HCC cells were co-injected with miR-320a over-expressing CAFs into nude mice. Besides EVs derived from CAFs, innate immune cell-derived MVs have also been reported to enhance HCC metastasis through CD11b and CD18, also known as integrin αMβ2.[88]

3.2 Cholangiocarcinoma

The presence of EVs in bile and their role regulating cholangiocyte physiology was first described in murine models.[89] However, EVs also play a role in biliary pathobiology. In CCA tumors, several reports have emphasized the importance of EVs in the regulation of the interplay between CCA cells and the cells present in the tumor stroma. CCA cell-derived EVs favour the fibroblastic differentiation of bone marrow-derived mesenchymal stem cells (MSCs) and the secretion of pro-inflammatory cytokines and chemokines, including interleukin (IL)-6, chemokine (C-X-C motif) ligand 1 (CLXC1), and chemokine (C-C motif) ligand 2 (CCL2/MCP-1), which ultimately stimulate CCA cell proliferation via IL6/STAT3 signalling pathway.[31]

CCA-derived EVs may contain oncogenic biomolecules not only involved in modulating inflammatory and proliferative responses but also controlling migratory and metastatic processes. Two studies employing comparative proteomic approaches have explored the protein content of CCA-derived and cholangiocyte-derived EVs in vitro, identifying significant differences and a particular oncogenic protein profile related to proliferation and motility in cancer cell-derived EVs.[90,91] Differentially expressed proteins involved in cholangiocarcinogenesis included EGFR, Mucin-1, integrin β4 (ITGB4), and epithelial cell adhesion molecule (EPCAM).[90] EGFR participates in CCA progression, favouring

the dedifferentiation and invasiveness of tumor cells and represents a bad prognostic factor.[92,93] Similarly, Mucin-1 and EPCAM, which are also upregulated in CCA, correlate with poor outcome in patients with CCA.[94–96] Interestingly, ITGB4 has recently been described as an EV integrin that dictates future metastatic sites, contributing to preferential organotropism of tumor cells.[38] On the other hand, EVs secreted by liver-fluke associated CCA cells induce cholangiocyte proliferation [97] and invasion,[91] events that are associated with an enrichment of oncoproteins in EVs, including galectin-3 binding protein (LG3BP), prostaglandin F2 receptor negative regulator, 4F2 cell-surface antigen heavy chain (4F2hc), integrin-β1 and EPCAM.[91]

4. NON-INVASIVE BIOMARKERS

The presence of EVs in biological fluids and their diverse molecular cargo has recently placed EVs as a new source of non-invasive disease biomarkers. Indeed, potential biomarker candidates (miRNAs and proteins) have been described in serum- and bile-derived EVs for the diagnosis and/or the prognosis prediction of HB, HCC, and CCA (**Table 2.1**).

In HB patients, serum EV miR-21 levels were higher than in healthy children, and negatively correlated with patient survival.[103] On the other hand, decreased levels of miR-34a, miR-34b, and miR-34c were reported in serum EVs from HB infants compared to healthy individuals. Combination of these miRs showed higher diagnostic value than the gold standard alpha fetoprotein (AFP).[104] Furthermore, reduced levels of the miR-34 panel in EVs of HB were associated with lower overall survival.[104]

In HCC patients, levels of miRs 18a, 221, 222, and 224 in serum EVs were found upregulated compared to patients with chronic hepatitis B (CHB) or liver cirrhosis, patients, whereas miR-101 level was downregulated.[101] Likewise, increased expression of miR-21 was identified in serum EVs from patients with HCC compared to CHB patients or healthy individuals, and correlated with cirrhosis and advanced tumor stage.[102] MiR-665 in serum EVs may also be a potential prognostic biomarker for HCC, as high miR-665 levels positively correlated with larger tumor size, local invasion and advanced clinical stages (stage III/IV), and negatively with overall survival.[99] Moreover, diminished levels of several miRNAs in serum EVs have been suggested as predictors of HCC recurrence or overall survival.[105,106] MiR expression profiling in serum EVs identified the tumor suppressor miR-718 downregulated in patients with larger tumor diameters and recurrence. Reduced miR-718 expression also correlated with poor histological tumor

cell differentiation.[105] Furthermore, low levels of miR-125b in serum EVs have been linked to advanced TNM stages and encapsulation, suggesting this miR as a potential prognostic candidate of recurrence and overall survival.[106] Besides miRNAs, different proteins present in serum EVs such as LG3BP, polymeric immunoglobulin receptor (PIGR) and alpha-2-macroglobulin (A2MG) were found upregulated in HCC patients compared to healthy individuals, with a better diagnostic value than AFP.[90] Apart from changes in the EV cargo, the EV concentration itself could also serve as a disease biomarker. In fact, stage I and II HCC patients showed higher EV concentration in serum compared to patients with liver cirrhosis.[107]

Table 2.1: EVs as non-invasive biomarkers of hepatobiliary malignancies.

DISEASE	NAME	BIOMARKER TYPE	EV SOURCE	NUMBER OF PATIENTS	EXPRESSION	SEN (%)	SPE (%)	AUC	REFERENCE
HB	miR-21	miRNA	Serum	HB (n=32) *vs* Control (n=32)	Up	—	—	0.861	96
	miR-34a* miR-34b* miR-34c*		Serum	HB (n=63) *vs* Control (n=63)	Down	— — —	— — —	0.963	97
HCC	LG3BP PIGR A2MG	Protein	Serum	HCC (n=29) *vs* Control (n=32)	Up	96.6 82.8 92.9	71.8 71.8 56.2	0.904 0.837 0.796	88
	MV (ug/mL)	Microvesicle concentration	Blood	Stage I HCC (n=28) *vs* Cirrhosis (n=40)	Up	—	—	0.83	98
				Stage II HCC (n=20) *vs* Cirrhosis (n=40)	Up	—	—	0.94	
	AnnexinV$^+$ EpCAM$^+$ (microparticle/mL)	TAMP concentration	Serum	HCC (n=86) *vs* Control (n=58)	Up	—	—	0.77	99
	AnnexinV$^+$ EpCAM$^+$ ASGPR1$^+$ (microparticle/mL)	TAMP concentration	Serum	HCC (n=86) *vs* Cirrhosis (n=49)	Up	—	—	0.73	
CCA	FIBG A1AG1 VTDB	Protein	Serum	iCCA (n=12) *vs* HCC (n=29)	Up	83.3 83.3 75.0	89.6 82.1 89.2	0.894 0.845 0.823	88
	AMPN VNN1 PIGR		Serum	CCA (n=43) *vs* Control (n=32)	Up	90.7 72.1 83.7	65.6 87.5 71.8	0.878 0.876 0.844	
	PIGR AMPN FIBG		Serum	CCA I-II (n=13) *vs* Control (n=22)	Up	75.0 91.7 100.0	95.4 72.7 68.1	0.905 0.833 0.833	
	FIBG A1AG1 S10A8		Serum	CCA (n=43) *vs* PSC (n=30)	Up	88.4 76.7 69.8	63.3 70.0 66.6	0.796 0.794 0.759	
	FCN2 ITIH4 FIBG		Serum	CCA I-II (n=13) *vs* PSC (n=30)	Up	100.0 91.7 91.7	80.9 80.9 80.9	0.956 0.881 0.881	
	miR-191* miR-486-3p* miR-1274b* miR-16* miR-484*	miRNA	Bile	CCA (n=46) *vs* Control (n=50; including PSC, biliary obstruction and bile leak)	Up	67.0	96.0	—	100
	ENST00000588480.1 ENST00000517758.1	lncRNA	Bile	CCA (n=35) *vs* Control (n=56)	Up	82.9	58.9	0.709	101
	nanoparticles/L	EV concentration	Bile	CBD stenoses (pancreatic cancer n=10 and CCA n=5) *vs* nonmalignant CBD stenoses (n=15)	Up	—	—	1.000	102
	AnnexinV$^+$ EpCAM$^+$ ASGPR1$^+$ (microparticle/mL)	TAMP concentration	Serum	CCA (n=38) *vs* Cirrhosis (n=49)	Up	—	-	0.63	99
Liver cancer (HCC/CCA)	AnnexinV$^+$ EpCAM$^+$ ASGPR1$^+$ (microparticle/mL)	TAMP concentration	Serum	Liver tumor (HCC n=86 and CCA n=38) *vs* Cirrhosis (n=49)	up	—	—	0.7	99

In CCA, a panel of miRs (191, 486-3p, 1274b, 16, 484) was found upregulated in bile EVs of patients with CCA compared to a control group containing PSC, biliary obstruction and bile leak syndrome patients.[108] The analysis of the lncRNA profile in bile EVs from CCA patients *vs.* patients with biliary obstruction identified the upregulation of two lncRNAs (*i.e.*, ENST00000588480.1 and ENST00000517758) in CCA patients.[98] The combined expression of both lncRNAs showed relevant diagnostic and prognostic value, being increased in advanced TNM stages (III-IV) and showing worse overall survival at high lncRNA concentrations. On the other hand, different proteins present in serum EVs exhibited high diagnostic values when comparing CCA patients with healthy individuals, such as aminopeptidase N (AMPN), pantetheinase (VNN1), and PIGR.[90] Some proteins present in serum EVs, such as ficolin-2 (FCN2), inter-alpha-trypsin inhibitor heavy chain H4 (ITIH4) and fibrinogen gamma chain (FIBG), displayed better diagnostic values than CA19-9 (a non-specific tumor marker for the diagnosis of CCA) in the differential diagnosis between CCA (stage I-II) and PSC.[90] Nowadays, the differential diagnosis between intrahepatic CCA (iCCA) and HCC by non-invasive methods is not feasible and compromises adequate treatment. In this regard, proteins present in serum EVs—such as FIBG, alpha-1-acid glycoprotein 1 (A1AG1) and vitamin-D binding protein (VTDB)—exhibited higher accuracy than CA19-9 and AFP for the differential diagnosis of iCCA *vs*. HCC.[90] As aforementioned, the EV concentration analysis could also be relevant for the diagnosis of malignant biliary diseases. In this regard, bile EV concentration was reported to accurately discriminate between malignant common bile duct (CBD) stenosis and nonmalignant CBD stenosis.[100] In addition, elevated concentration of AnnexinV/EpCAM/ASGPR1 positive tumor-associated microparticles (TAMPs) allowed the diagnosis of patients with liver cancer (HCC and CCA) compared to cirrhotic patients, while no changes were detected between HCC and CCA.[108] Notably, the levels of these TAMPs decreased 7 days after the surgical resection of liver tumors, closely relating this microparticle population with tumor presence.

5. THERAPEUTIC IMPLICATIONS

The use of EVs in anti-cancer therapy is currently under investigation. As EVs carry different types of molecules, they can be used as vehicles to deliver therapeutic cargo into cancer cells.[109] Moreover, EVs have shown the ability to modulate the immune system, and to stimulate the immune response against tumor cells.[110]

5.1 Molecule Carriers

EVs as therapeutic delivery systems provide benefits for the carried therapeutic molecule. Hence, encapsulation of therapeutic compounds (such as chemicals, RNAs, DNAs, proteins, or lipids) increases their bioavailability by preserving their integrity and biological activity, as well as protecting them from enzymatic degradation in biological fluids.[111] In comparison to other therapeutic vectors such as synthetic nano-particles, liposomes or recombinant viral vectors, EVs are generally non-immunogenic in nature, which enhances their resistance to fast clearance from circulation.[111] EVs also display low toxicity and are quite stable in tissues and circulation, representing adequate therapeutic delivery systems against cancer.[112] Furthermore, cell type-specific proteins within EVs seem to provide certain cell tropism.[111]

The strategy of using EVs as therapeutic molecule delivery vehicles is starting in liver cancer, mainly focusing on miRNAs. Stellate cell-derived EVs loaded with miR-335-5p, a tumor suppressor miR downregulated in HCC, inhibits HCC cell invasiveness in vitro and induces HCC tumor shrinkage in vivo through the repression of proliferation and stimulation of apoptosis.[113] Moreover, miR-122 enriched EVs obtained from adipose tissue-derived mesenchymal stem cells (ADMSCs) increases HCC cell sensitivity to the chemotherapeutic agents sorafenib and 5-FU.[114] The underlying mechanism regulating chemosensitivity consists on the downregulation of miR-122 target genes including cyclin G1 (CCNG1), disintegrin and metalloproteinase domain-containing protein 10 (ADAM10), and insulin-like growth factor 1 receptor (IGF1R), which induce apoptosis and cell cycle arrest in vitro. Furthermore, intra-tumor injection of miR-122-enriched EVs in a HCC xenograft mouse model synergized the inhibitory effect of sorafenib in vivo, reducing tumor size.[114]

In CCA, stellate cell-derived EVs carrying miR-195 inhibited CCA growth and invasiveness in vitro.[115] Tail vein injection of miR-195 loaded EVs into an orthotopic rat model of CCA reduced tumor size and improved the overall animal survival.[115] These anti-neoplasic effects are likely mediated via targeting VEGF, cell division control (CDC) proteins 25 and 42, as well as cyclin-dependent kinases (CDK) 1, 4, and 6.

5.2 Immunotherapy

An alternative therapeutic strategy contemplates the use of EVs as stimulators of the immune system in order to elicit a nontoxic, systemic, and long-lived anti-tumor immune

response. Different studies have described the potential use of EVs as immunostimulatory entities against HCC.[116–120] For instance, HCC cells under stress conditions, such as heat shock or chemotherapeutic anti-cancer drug treatment, increased EV secretion and surface expression of heat shock proteins (HSPs).[116] HSP-bearing EVs can boost natural killer (NK) cell-mediated cytotoxic response against HCC cells in vitro.[116] Similarly, histone deacetylase inhibitor MS-275 enhanced the protein levels of immunostimulatory molecules [MHC class I polypeptide-related sequence B (MICB) and HSP70] in EVs derived from HCC cells, increasing the cytotoxicity of NK cells and anti-tumor response.[117] The anti-HCC tumor immune response can also be induced by ADMSC-derived EVs, which promote natural killer T cell (NKT) anti-tumor response, thereby facilitating HCC suppression.[118]

Alternatively, HCC cell-derived EVs display HCC antigens AFP and glypican 3. Capture of these EVs by dendritic cells (DCs) triggers a strong DC-mediated T cell dependent anti-tumor immune response both in vitro and in ectopic and orthotopic in vivo mouse models.[119] EVs from antigen presenting cells (APCs) can also induce anti-tumor immune responses against HCC. EVs derived from AFP-expressing DCs are able to trigger potent antigen-specific anti-tumor immune responses and reshape the tumor microenvironment from an immunoinhibitory to an immunostimulatory setting in diverse HCC mice models including ectopic, orthotopic and carcinogen-induced HCC.[120] Thus, AFP-expressing DC-derived EVs stimulate antigen-specific anti-tumor immune responses in vivo, eliciting suppression of tumor growth and prolonging mice survival.[120]

6. CONCLUDING REMARKS AND FUTURE DIRECTIONS

Early diagnosis and treatment of hepatobiliary malignancies is still far from being manageable. The development of non-invasive diagnostic and disease monitoring tools represents a major challenge. The presence of EVs in biological fluids, as well as their capacity to carry tumor-associated molecules, make EVs excellent candidates for clinical application. Hereof, certain progress is being made in the potential use of EVs as a source of non-invasive disease biomarkers. EV concentration as well as their specific cargo can serve as indicators of the different pathological stages of a disease, including the discrimination between early and late phases, and estimation of recurrence and metastasis risk. For that matter, the application of omic technologies has provided some potential candidate biomarkers. However, in order to transfer knowledge into the clinical practice, several limitations, and concerns should be considered: (i) different EV isolation

procedures (*i.e.*, ultracentrifugation, size exclusion, immune-affinity isolation, polymeric precipitation, and microfluidics) are currently used, providing diverse EV populations and yield depending on the nature of the isolation protocol (ii) a proper characterization of the EVs fraction should be performed. There are minimal experimental requirements defined by the International Society for Extracellular Vesicles (ISEV),[121] which include the analysis of the EV quantity [*e.g.*, nanoparticle tracking analysis (NTA), IZON qNano technique, flow cytometry], size [*e.g.*, NTA, IZON qNano technique, electron microscopy, dynamic light scattering (DLS)], and presence of specific surface markers (*e.g.*, immunoblot, immune-gold electron microscopy),[121,122] (iii) specific EV markers to distinguish EV subpopulations according to their origin (*e.g.*, exosomes, MVs, apoptotic bodies) are still missing,[121] (iv) appropriate clinically-relevant control groups with biopsy-proven diagnosis, as well as a representative number of samples should be included to ensure the accuracy (sensitivity, specificity, AUC, predictive and likelihood ratio values) and significance of the results,[123] (v) candidate biomarkers identified in a discovery phase must be internationally validated using easily transferable methodologies into the clinical settings (*e.g.*, ELISA, qPCR), ideally using raw biological fluids (*i.e.*, serum, urine, saliva) and avoiding the costly and time consuming EV isolation techniques.

EVs represent a new opportunity for cancer therapy. They participate in the development and progression of cancer, including the formation of a pro-tumorigenic microenvironment, angiogenesis, chemoresistance, and the generation of a metastatic niche, promoting tumor growth, and aggressiveness. Therefore, interfering the EV biogenesis and/or release may be a potential therapeutic strategy. Several inhibitors targeting these crucial steps have been developed (**Figure 2.1**), although their safety and efficacy should be clinically evaluated in the future. Nevertheless, additional regulatory mechanisms of EV generation (*e.g.*, loading), trafficking and autocrine/paracrine signal transduction (*e.g.*, recipient cell internalization routes of specific EV subpopulations) need to be elucidated, which could provide other targets for therapy.[124] On the other hand, EVs could be used as drug delivery systems and as immunomodulators promoting anti-tumor response. For drug delivery, a major challenge represents the specific cell targeting in vivo, as well as the use of immunologically inert and biocompatible EVs. In contrast, the capacity of EVs to regulate the immune system opens new opportunities for targeting malignancies and for developing anti-tumor vaccines.[125]

In conclusion, EVs represent an emerging and stimulating field of research in liver cancer with multiple potential applications, from biomarker discovery to therapy. Nonetheless,

thorough research is still needed to gain knowledge on their intrinsic role in liver health and disease, and to evaluate their potential clinical application.

REFERENCES

1. Sia D, Villanueva A, Friedman SL, Llovet JM. Liver Cancer Cell of Origin, Molecular Class, and Effects on Patient Prognosis. *Gastroenterology*. 2017;152(4):745-761. doi:10.1053/J.GASTRO.2016.11.048

2. Rizvi S, Gores GJ. Pathogenesis, diagnosis, and management of cholangiocarcinoma. *Gastroenterology*. 2013;145(6):1215-1229. doi:10.1053/J.GASTRO.2013.10.013

3. Banales JM, Cardinale V, Carpino G, et al. Expert consensus document: Cholangiocarcinoma: current knowledge and future perspectives consensus statement from the European Network for the Study of Cholangiocarcinoma (ENS-CCA). *Nat Rev Gastroenterol Hepatol*. 2016;13(5):261-280. doi:10.1038/NRGASTRO.2016.51

4. Czauderna P, Lopez-Terrada D, Hiyama E, Häberle B, Malogolowkin MH, Meyers RL. Hepatoblastoma state of the art: pathology, genetics, risk stratification, and chemotherapy. *Curr Opin Pediatr*. 2014;26(1):19-28. doi:10.1097/MOP.0000000000000046

5. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet (London, England)*. 2018;391(10127):1301-1314. doi:10.1016/S0140-6736(18)30010-2

6. Bruix J, Gores GJ, Mazzaferro V. Hepatocellular carcinoma: clinical frontiers and perspectives. *Gut*. 2014;63(5):844-855. doi:10.1136/GUTJNL-2013-306627

7. Schnater JM, Köhler SE, Lamers WH, Von Schweinitz D, Aronson DC. Where do we stand with hepatoblastoma? A review. *Cancer*. 2003;98(4):668-678. doi:10.1002/CNCR.11585

8. Kahlert C, Melo SA, Protopopov A, et al. Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer. *J Biol Chem*. 2014;289(7):3869-3875. doi:10.1074/JBC.C113.532267

9. Melo SA, Luecke LB, Kahlert C, et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature*. 2015;523(7559):177-182. doi:10.1038/NATURE14581

10. Yáñez-Mó M, Siljander PRM, Andreu Z, et al. Biological properties of extracellular vesicles and their physiological functions. *J Extracell vesicles*. 2015;4(2015):1-60. doi:10.3402/JEV.V4.27066

11. Raposo G, Stoorvogel W. Extracellular vesicles: exosomes, microvesicles, and friends. *J Cell Biol*. 2013;200(4):373-383. doi:10.1083/JCB.201211138

12. Colombo M, Raposo G, Théry C. Biogenesis, secretion, and intercellular interactions of exosomes and other extracellular vesicles. *Annu Rev Cell Dev Biol*. 2014;30:255-289. doi:10.1146/ANNUREV-CELLBIO-101512-122326

13. Tkach M, Théry C. Communication by Extracellular Vesicles: Where We Are and Where We Need to Go. *Cell*. 2016;164(6):1226-1232. doi:10.1016/J.CELL.2016.01.043

14. Huotari J, Helenius A. Endosome maturation. *EMBO J*. 2011;30(17):3481-3500. doi:10.1038/EMBOJ.2011.286

15. Akers JC, Gonda D, Kim R, Carter BS, Chen CC. Biogenesis of extracellular vesicles (EV): exosomes, microvesicles, retrovirus-like vesicles, and apoptotic bodies. *J Neurooncol*. 2013;113(1):1-11. doi:10.1007/S11060-013-1084-8

16. György B, Szabó TG, Pásztói M, et al. Membrane vesicles, current state-of-the-art: emerging role of extracellular vesicles. *Cell Mol Life Sci*. 2011;68(16):2667-2688. doi:10.1007/S00018-011-0689-3

17. Ariazi J, Benowitz A, De Biasi V, et al. Tunneling Nanotubes and Gap Junctions-Their Role in Long-Range Intercellular Communication during Development, Health, and Disease Conditions. *Front Mol Neurosci*. 2017;10. doi:10.3389/FNMOL.2017.00333

18. Ramachandran S, Palanisamy V. Horizontal transfer of RNAs: exosomes as mediators of intercellular communication. *Wiley Interdiscip Rev RNA*. 2012;3(2):286-293. doi:10.1002/WRNA.115

19. Liem M, Ang CS, Mathivanan S. Insulin Mediated Activation of PI3K/Akt Signalling Pathway Modifies the Proteomic Cargo of Extracellular Vesicles. *Proteomics*. 2017;17(23-24). doi:10.1002/PMIC.201600371

20. Saha B, Momen-Heravi F, Kodys K, Szabo G. MicroRNA Cargo of Extracellular Vesicles from Alcohol-exposed Monocytes Signals Naive Monocytes to Differentiate into M2 Macrophages. *J Biol Chem*. 2016;291(1):149-159. doi:10.1074/JBC.M115.694133

21. Keller S, Ridinger J, Rupp AK, Janssen JWG, Altevogt P. Body fluid derived exosomes as a novel template for clinical diagnostics. *J Transl Med*. 2011;9. doi:10.1186/1479-5876-9-86

22. Ogorevc E, Kralj-Iglic V, Veranic P. The role of extracellular vesicles in phenotypic cancer transformation. *Radiol Oncol*. 2013;47(3):197-205. doi:10.2478/RAON-2013-0037

23. Lindoso RS, Collino F, Vieyra A. Extracellular vesicles as regulators of tumor fate: crosstalk among cancer stem cells, tumor cells and mesenchymal stem cells. *Stem cell Investig*. 2017;4(9). doi:10.21037/SCI.2017.08.08

24. Katsuda T, Kosaka N, Ochiya T. The roles of extracellular vesicles in cancer biology: toward the development of novel cancer biomarkers. *Proteomics*. 2014;14(4-5):412-425. doi:10.1002/PMIC.201300389

25. Sadovska L, Santos CB, Kalniņa Z, Linē A. Biodistribution, Uptake and Effects Caused by Cancer-Derived Extracellular Vesicles. *J Circ biomarkers*. 2015;4. doi:10.5772/60522

26. Kogure T, Lin WL, Yan IK, Braconi C, Patel T. Intercellular nanovesicle-mediated microRNA transfer: a mechanism of environmental modulation of hepatocellular cancer cell growth. *Hepatology*. 2011;54(4):1237-1248. doi:10.1002/HEP.24504

27. Skog J, Würdinger T, van Rijn S, et al. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat Cell Biol*. 2008;10(12):1470-1476. doi:10.1038/NCB1800

28. Khan S, Aspe JR, Asumen MG, et al. Extracellular, cell-permeable survivin inhibits apoptosis while promoting proliferative and metastatic potential. *Br J Cancer*. 2009;100(7):1073-1086. doi:10.1038/SJ.BJC.6604978

29. Xiao D, Barry S, Kmetz D, et al. Melanoma cell-derived exosomes promote epithelial-mesenchymal transition in primary melanocytes through paracrine/autocrine signaling in the tumor microenvironment. *Cancer Lett*. 2016;376(2):318-327. doi:10.1016/J.CANLET.2016.03.050

30. Harris DA, Patel SH, Gucek M, Hendrix A, Westbroek W, Taraska JW. Exosomes released from breast cancer carcinomas stimulate cell movement. *PLoS One*. 2015;10(3). doi:10.1371/JOURNAL.PONE.0117495

31. Haga H, Yan IK, Takahashi K, Wood J, Zubair A, Patel T. Tumour cell-derived extracellular vesicles interact with mesenchymal stem cells to modulate the microenvironment and enhance cholangiocarcinoma growth. *J Extracell vesicles*. 2015;4(2015):1-13. doi:10.3402/JEV.V4.24900

32. Mu W, Rana S, Zöller M. Host matrix modulation by tumor exosomes promotes motility and invasiveness. *Neoplasia*. 2013;15(8):875-887. doi:10.1593/NEO.13786

33. Bouvy C, Gheldof D, Chatelain C, Mullier F, Dogné JM. Contributing role of extracellular vesicles on vascular endothelium haemostatic balance in cancer. *J Extracell vesicles*. 2014;3(1). doi:10.3402/JEV.V3.24400

34. Grange C, Tapparo M, Collino F, et al. Microvesicles released from human renal cancer stem cells stimulate angiogenesis and formation of lung premetastatic niche. *Cancer Res*. 2011;71(15):5346-5356. doi:10.1158/0008-5472.CAN-11-0241

35. Shao H, Chung J, Lee K, et al. Chip-based analysis of exosomal mRNA mediating drug resistance in glioblastoma. *Nat Commun*. 2015;6. doi:10.1038/NCOMMS7999

36. Corcoran C, Rani S, O'Brien K, et al. Docetaxel-resistance in prostate cancer: evaluating associated phenotypic changes and potential for resistance transfer via exosomes. *PLoS One*. 2012;7(12). doi:10.1371/JOURNAL.PONE.0050999

37. Chen WX, Liu XM, Lv MM, et al. Exosomes from drug-resistant breast cancer cells transmit chemoresistance by a horizontal transfer of microRNAs. *PLoS One*. 2014;9(4). doi:10.1371/JOURNAL.PONE.0095240

38. Hoshino A, Costa-Silva B, Shen TL, et al. Tumour exosome integrins determine organotropic metastasis. *Nature*. 2015;527(7578):329-335. doi:10.1038/NATURE15756

39. Peinado H, Alečković M, Lavotshkin S, et al. Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. *Nat Med*. 2012;18(6):883-891. doi:10.1038/NM.2753

40. Chalmin F, Ladoire S, Mignot G, et al. Membrane-associated Hsp72 from tumor-derived exosomes mediates STAT3-dependent immunosuppressive function of mouse and human myeloid-derived suppressor cells. *J Clin Invest*. 2010;120(2):457-471. doi:10.1172/JCI40483

41. Valenti R, Huber V, Filipazzi P, et al. Human tumor-released microvesicles promote the differentiation of myeloid cells with transforming growth factor-beta-mediated suppressive activity on T lymphocytes. *Cancer Res*. 2006;66(18):9290-9298. doi:10.1158/0008-5472.CAN-06-1819

42. Klibi J, Niki T, Riedel A, et al. Blood diffusion and Th1-suppressive effects of galectin-9-containing exosomes released by Epstein-Barr virus-infected nasopharyngeal carcinoma cells. *Blood*. 2009;113(9):1957-1966. doi:10.1182/BLOOD-2008-02-142596

43. Van Niel G, D'Angelo G, Raposo G. Shedding light on the cell biology of extracellular vesicles. *Nat Rev Mol Cell Biol*. 2018;19(4):213-228. doi:10.1038/NRM.2017.125

44. Colombo M, Moita C, Van Niel G, et al. Analysis of ESCRT functions in exosome biogenesis, composition and secretion highlights the heterogeneity of extracellular vesicles. *J Cell Sci*. 2013;126(Pt 24):5553-5565. doi:10.1242/JCS.128868

45. Trajkovic K, Hsu C, Chiantia S, et al. Ceramide triggers budding of exosome vesicles into multivesicular endosomes. *Science*. 2008;319(5867):1244-1247. doi:10.1126/SCIENCE.1153124

46. Gulbins E, Kolesnick R. Raft ceramide in molecular medicine. *Oncogene*. 2003;22(45):7070-7077. doi:10.1038/SJ.ONC.1207146

47. De Gassart A, Géminard C, Février B, Raposo G, Vidal M. Lipid raft-associated protein sorting in exosomes. *Blood*. 2003;102(13):4336-4344. doi:10.1182/BLOOD-2003-03-0871

48. Laulagnier K, Grand D, Dujardin A, et al. PLD2 is enriched on exosomes and its activity is correlated to the release of exosomes. *FEBS Lett*. 2004;572(1-3):11-14. doi:10.1016/J.FEBSLET.2004.06.082

49. Ghossoub R, Lembo F, Rubio A, et al. Syntenin-ALIX exosome biogenesis and budding into multivesicular bodies are controlled by ARF6 and PLD2. *Nat Commun*. 2014;5. doi:10.1038/NCOMMS4477

50. Perez-Hernandez D, Gutiérrez-Vázquez C, Jorge I, et al. The intracellular interactome of tetraspanin-enriched microdomains reveals their function as sorting machineries toward exosomes. *J Biol Chem*. 2013;288(17):11649-11661. doi:10.1074/JBC.M112.445304

51. van Niel G, Charrin S, Simoes S, et al. The tetraspanin CD63 regulates ESCRT-independent and -dependent endosomal sorting during melanogenesis. *Dev Cell*. 2011;21(4):708-721. doi:10.1016/J.DEVCEL.2011.08.019

52. Xiang X, Qiu R, Yao X, Arst HN, Peñalva MA, Zhang J. Cytoplasmic dynein and early endosome transport. *Cell Mol Life Sci*. 2015;72(17):3267-3280. doi:10.1007/S00018-015-1926-Y

53. Bonifacino JS, Glick BS. The mechanisms of vesicle budding and fusion. *Cell*. 2004;116(2):153-166. doi:10.1016/S0092-8674(03)01079-1

54. Jahn R, Scheller RH. SNAREs--engines for membrane fusion. *Nat Rev Mol Cell Biol*. 2006;7(9):631-643. doi:10.1038/NRM2002

55. Piccin A, Murphy WG, Smith OP. Circulating microparticles: pathophysiology and clinical implications. *Blood Rev*. 2007;21(3):157-171. doi:10.1016/J.BLRE.2006.09.001

56. Mulcahy LA, Pink RC, Carter DRF. Routes and mechanisms of extracellular vesicle uptake. *J Extracell vesicles*. 2014;3(1). doi:10.3402/JEV.V3.24641

57. Singh R, Pochampally R, Watabe K, Lu Z, Mo YY. Exosome-mediated transfer of miR-10b promotes cell invasion in breast cancer. *Mol Cancer*. 2014;13(1). doi:10.1186/1476-4598-13-256

58. Richards KE, Zeleniak AE, Fishel ML, Wu J, Littlepage LE, Hill R. Cancer-associated fibroblast exosomes regulate survival and proliferation of pancreatic cancer cells. *Oncogene*. 2017;36(13):1770-1778. doi:10.1038/ONC.2016.353

59. Kowal J, Tkach M, Théry C. Biogenesis and secretion of exosomes. *Curr Opin Cell Biol*. 2014;29(1):116-125. doi:10.1016/J.CEB.2014.05.004

60. Bobrie A, Krumeich S, Reyal F, et al. Rab27a supports exosome-dependent and -independent mechanisms that modify the tumor microenvironment and can promote tumor progression. *Cancer Res*. 2012;72(19):4920-4930. doi:10.1158/0008-5472.CAN-12-0925

61. Escola JM, Kleijmeer MJ, Stoorvogel W, Griffith JM, Yoshie O, Geuze HJ. Selective

enrichment of tetraspan proteins on the internal vesicles of multivesicular endosomes and on exosomes secreted by human B-lymphocytes. *J Biol Chem*. 1998;273(32):20121-20127. doi:10.1074/JBC.273.32.20121

62. Théry C, Regnault A, Garin J, et al. Molecular characterization of dendritic cell-derived exosomes. Selective accumulation of the heat shock protein hsc73. *J Cell Biol*. 1999;147(3):599-610. doi:10.1083/JCB.147.3.599

63. Morelli AE, Larregina AT, Shufesky WJ, et al. Endocytosis, intracellular sorting, and processing of exosomes by dendritic cells. *Blood*. 2004;104(10):3257-3266. doi:10.1182/BLOOD-2004-03-0824

64. Hemler ME. Tetraspanin functions and associated microdomains. *Nat Rev Mol Cell Biol*. 2005;6(10):801-811. doi:10.1038/NRM1736

65. Barrès C, Blanc L, Bette-Bobillo P, et al. Galectin-5 is bound onto the surface of rat reticulocyte exosomes and modulates vesicle uptake by macrophages. *Blood*. 2010;115(3):696-705. doi:10.1182/BLOOD-2009-07-231449

66. Näslund TI, Paquin-Proulx D, Paredes PT, Vallhov H, Sandberg JK, Gabrielsson S. Exosomes from breast milk inhibit HIV-1 infection of dendritic cells and subsequent viral transfer to CD4+ T cells. *AIDS*. 2014;28(2):171-180. doi:10.1097/QAD.0000000000000159

67. Hao S, Bai O, Li F, Yuan J, Laferte S, Xiang J. Mature dendritic cells pulsed with exosomes stimulate efficient cytotoxic T-lymphocyte responses and antitumour immunity. *Immunology*. 2007;120(1):90-102. doi:10.1111/J.1365-2567.2006.02483.X

68. Christianson HC, Svensson KJ, Van Kuppevelt TH, Li JP, Belting M. Cancer cell exosomes depend on cell-surface heparan sulfate proteoglycans for their internalization and functional activity. *Proc Natl Acad Sci U S A*. 2013;110(43):17380-17385. doi:10.1073/PNAS.1304266110/-/DCSUPPLEMENTAL/PNAS.201304266SI.PDF

69. Atai NA, Balaj L, Van Veen H, et al. Heparin blocks transfer of extracellular vesicles between donor and recipient cells. *J Neurooncol*. 2013;115(3):343-351. doi:10.1007/S11060-013-1235-Y

70. Escrevente C, Keller S, Altevogt P, Costa J. Interaction and uptake of exosomes by ovarian cancer cells. *BMC Cancer*. 2011;11. doi:10.1186/1471-2407-11-108

71. Montecalvo A, Larregina AT, Shufesky WJ, et al. Mechanism of transfer of functional microRNAs between mouse dendritic cells via exosomes. *Blood*. 2012;119(3):756-766. doi:10.1182/BLOOD-2011-02-338004

72. Feng D, Zhao WL, Ye YY, et al. Cellular internalization of exosomes occurs through phagocytosis. *Traffic*. 2010;11(5):675-687. doi:10.1111/J.1600-0854.2010.01041.X

73. Herskovits JS, Burgess CC, Obar RA, Vallee RB. Effects of mutant rat dynamin on endocytosis. *J Cell Biol*. 1993;122(3):565-578. doi:10.1083/JCB.122.3.565

74. Vallee RB, Herskovits JS, Aghajanian JG, Burgess CC, Shpetner HS. Dynamin, a GTPase involved in the initial stages of endocytosis. *Ciba Found Symp*. 1993;176. doi:10.1002/9780470514450.CH12

75. Ehrlich M, Boll W, Van Oijen A, et al. Endocytosis by random initiation and stabilization of clathrin-coated pits. *Cell*. 2004;118(5):591-605. doi:10.1016/J.CELL.2004.08.017

76. Fitzner D, Schnaars M, Van Rossum D, et al. Selective transfer of exosomes from oligodendrocytes to microglia by macropinocytosis. *J Cell Sci*. 2011;124(Pt 3):447-458.

doi:10.1242/JCS.074088

77. Izquierdo-Useros N, Naranjo-Gómez M, Archer J, et al. Capture and transfer of HIV-1 particles by mature dendritic cells converges with the exosome-dissemination pathway. *Blood*. 2009;113(12):2732-2741. doi:10.1182/BLOOD-2008-05-158642

78. Koumangoye RB, Sakwe AM, Goodwin JS, Patel T, Ochieng J. Detachment of breast tumor cells induces rapid secretion of exosomes which subsequently mediate cellular adhesion and spreading. *PLoS One*. 2011;6(9). doi:10.1371/JOURNAL.PONE.0024234

79. Svensson KJ, Christianson HC, Wittrup A, et al. Exosome uptake depends on ERK1/2-heat shock protein 27 signaling and lipid Raft-mediated endocytosis negatively regulated by caveolin-1. *J Biol Chem*. 2013;288(24):17713-17724. doi:10.1074/JBC.M112.445403

80. Al-Nedawi K, Meehan B, Kerbel RS, Allison AC, Rak A. Endothelial expression of autocrine VEGF upon the uptake of tumor-derived microvesicles containing oncogenic EGFR. *Proc Natl Acad Sci U S A*. 2009;106(10):3794-3799. doi:10.1073/PNAS.0804543106

81. Corbet C, Feron O. Tumour acidosis: from the passenger to the driver's seat. *Nat Rev Cancer*. 2017;17(10):577-593. doi:10.1038/NRC.2017.77

82. Parolini I, Federici C, Raggi C, et al. Microenvironmental pH is a key factor for exosome traffic in tumor cells. *J Biol Chem*. 2009;284(49):34211-34222. doi:10.1074/JBC.M109.041152

83. Takahashi K, Yan IK, Kogure T, Haga H, Patel T. Extracellular vesicle-mediated transfer of long non-coding RNA ROR modulates chemosensitivity in human hepatocellular cancer. *FEBS Open Bio*. 2014;4:458-467. doi:10.1016/J.FOB.2014.04.007

84. Qu Z, Wu J, Wu J, Luo D, Jiang C, Ding Y. Exosomes derived from HCC cells induce sorafenib resistance in hepatocellular carcinoma both in vivo and in vitro. *J Exp Clin Cancer Res*. 2016;35(1):1-12. doi:10.1186/S13046-016-0430-Z

85. Conigliaro A, Costa V, Lo Dico A, et al. CD90+ liver cancer cells modulate endothelial cell phenotype through the release of exosomes containing H19 lncRNA. *Mol Cancer*. 2015;14(1). doi:10.1186/S12943-015-0426-X

86. He M, Qin H, Poon TCW, et al. Hepatocellular carcinoma-derived exosomes promote motility of immortalized hepatocyte through transfer of oncogenic proteins and RNAs. *Carcinogenesis*. 2015;36(9):1008-1018. doi:10.1093/CARCIN/BGV081

87. Zhang Z, Li X, Sun W, et al. Loss of exosomal miR-320a from cancer-associated fibroblasts contributes to HCC proliferation and metastasis. *Cancer Lett*. 2017;397:33-42. doi:10.1016/J.CANLET.2017.03.004

88. Ma J, Cai W, Zhang Y, et al. Innate immune cell-derived microparticles facilitate hepatocarcinoma metastasis by transferring integrin $\alpha(M)\beta_2$ to tumor cells. *J Immunol*. 2013;191(6):3453-3461. doi:10.4049/JIMMUNOL.1300171

89. Masyuk AI, Huang BQ, Ward CJ, et al. Biliary exosomes influence cholangiocyte regulatory mechanisms and proliferation through interaction with primary cilia. *Am J Physiol Gastrointest Liver Physiol*. 2010;299(4). doi:10.1152/AJPGI.00093.2010

90. Arbelaiz A, Azkargorta M, Krawczyk M, et al. Serum extracellular vesicles contain protein biomarkers for primary sclerosing cholangitis and cholangiocarcinoma. *Hepatology*. 2017;66(4):1125-1143. doi:10.1002/HEP.29291

91. Dutta S, Reamtong O, Panvongsa W, et al. Proteomics profiling of cholangiocarcinoma exosomes: A potential role of oncogenic protein transferring in cancer progression. *Biochim Biophys Acta*. 2015;1852(9):1989-1999. doi:10.1016/J.BBADIS.2015.06.024

92. Clapéron A, Mergey M, Nguyen Ho-Bouldoires TH, et al. EGF/EGFR axis contributes to the progression of cholangiocarcinoma through the induction of an epithelial-mesenchymal transition. *J Hepatol*. 2014;61(2):325-332. doi:10.1016/J.JHEP.2014.03.033

93. Yoshikawa D, Ojima H, Iwasaki M, et al. Clinicopathological and prognostic significance of EGFR, VEGF, and HER2 expression in cholangiocarcinoma. *Br J Cancer*. 2008;98(2):418-425. doi:10.1038/SJ.BJC.6604129

94. Tamada S, Shibahara H, Higashi M, et al. MUC4 is a novel prognostic factor of extrahepatic bile duct carcinoma. *Clin Cancer Res*. 2006;12(14 Pt 1):4257-4264. doi:10.1158/1078-0432.CCR-05-2814

95. Sasaki M, Matsubara T, Yoneda N, et al. Overexpression of enhancer of zeste homolog 2 and MUC1 may be related to malignant behaviour in intraductal papillary neoplasm of the bile duct. *Histopathology*. 2013;62(3):446-457. doi:10.1111/HIS.12016

96. Sulpice L, Rayar M, Turlin B, et al. Epithelial cell adhesion molecule is a prognosis marker for intrahepatic cholangiocarcinoma. *J Surg Res*. 2014;192(1):117-123. doi:10.1016/J.JSS.2014.05.017

97. Chaiyadet S, Sotillo J, Smout M, et al. Carcinogenic Liver Fluke Secretes Extracellular Vesicles That Promote Cholangiocytes to Adopt a Tumorigenic Phenotype. *J Infect Dis*. 2015;212(10):1636-1645. doi:10.1093/INFDIS/JIV291

98. Ge X, Wang Y, Nie J, et al. The diagnostic/prognostic potential and molecular functions of long non-coding RNAs in the exosomes derived from the bile of human cholangiocarcinoma. *Oncotarget*. 2017;8(41):69995-70005. doi:10.18632/ONCOTARGET.19547

99. Qu Z, Wu J, Wu J, et al. Exosomal miR-665 as a novel minimally invasive biomarker for hepatocellular carcinoma diagnosis and prognosis. *Oncotarget*. 2017;8(46):80666-80678. doi:10.18632/ONCOTARGET.20881

100. Severino V, Dumonceau JM, Delhaye M, et al. Extracellular Vesicles in Bile as Markers of Malignant Biliary Stenoses. *Gastroenterology*. 2017;153(2):495-504.e8. doi:10.1053/J.GASTRO.2017.04.043

101. Sohn W, Kim J, Kang SH, et al. Serum exosomal microRNAs as novel biomarkers for hepatocellular carcinoma. *Exp Mol Med*. 2015;47(9). doi:10.1038/EMM.2015.68

102. Wang H, Hou L, Li A, Duan Y, Gao H, Song X. Expression of serum exosomal microRNA-21 in human hepatocellular carcinoma. *Biomed Res Int*. 2014;2014. doi:10.1155/2014/864894

103. Liu W, Chen S, Liu B. Diagnostic and prognostic values of serum exosomal microRNA-21 in children with hepatoblastoma: a Chinese population-based study. *Pediatr Surg Int*. 2016;32(11):1059-1065. doi:10.1007/S00383-016-3960-8

104. Jiao C, Jiao X, Zhu A, Ge J, Xu X. Exosomal miR-34s panel as potential novel diagnostic and prognostic biomarker in patients with hepatoblastoma. *J Pediatr Surg*. 2017;52(4):618-624. doi:10.1016/J.JPEDSURG.2016.09.070

105. Sugimachi K, Matsumura T, Hirata H, et al. Identification of a bona fide microRNA

biomarker in serum exosomes that predicts hepatocellular carcinoma recurrence after liver transplantation. *Br J Cancer*. 2015;112(3):532-538. doi:10.1038/BJC.2014.621

106. Liu W, Hu J, Zhou K, et al. Serum exosomal miR-125b is a novel prognostic marker for hepatocellular carcinoma. *Onco Targets Ther*. 2017;10:3843-3851. doi:10.2147/OTT.S140062

107. Wang W, Li H, Zhou Y, Jie S. Peripheral blood microvesicles are potential biomarkers for hepatocellular carcinoma. *Cancer Biomark*. 2013;13(5):351-357. doi:10.3233/CBM-130370

108. Julich-Haertel H, Urban SK, Krawczyk M, et al. Cancer-associated circulating large extracellular vesicles in cholangiocarcinoma and hepatocellular carcinoma. *J Hepatol*. 2017;67(2):282-292. doi:10.1016/J.JHEP.2017.02.024

109. György B, Hung ME, Breakefield XO, Leonard JN. Therapeutic applications of extracellular vesicles: clinical promise and open questions. *Annu Rev Pharmacol Toxicol*. 2015;55:439-464. doi:10.1146/ANNUREV-PHARMTOX-010814-124630

110. Robbins PD, Morelli AE. Regulation of immune responses by extracellular vesicles. *Nat Rev Immunol*. 2014;14(3):195-208. doi:10.1038/NRI3622

111. Kim SM, Kim HS. Engineering of extracellular vesicles as drug delivery vehicles. *Stem Cell Investig*. 2017;4(9). doi:10.21037/SCI.2017.08.07

112. Santangelo L, Battistelli C, Montaldo C, Citarella F, Strippoli R, Cicchini C. Functional Roles and Therapeutic Applications of Exosomes in Hepatocellular Carcinoma. *Biomed Res Int*. 2017;2017. doi:10.1155/2017/2931813

113. Wang F, Li L, Piontek K, Sakaguchi M, Selaru FM. Exosome miR-335 as a novel therapeutic strategy in hepatocellular carcinoma. *Hepatology*. 2018;67(3):940-954. doi:10.1002/HEP.29586

114. Lou G, Song X, Yang F, et al. Exosomes derived from miR-122-modified adipose tissue-derived MSCs increase chemosensitivity of hepatocellular carcinoma. *J Hematol Oncol*. 2015;8(1). doi:10.1186/S13045-015-0220-7

115. Li L, Piontek K, Ishida M, et al. Extracellular vesicles carry microRNA-195 to intrahepatic cholangiocarcinoma and improve survival in a rat model. *Hepatology*. 2017;65(2):501-514. doi:10.1002/HEP.28735

116. Lv LH, Wan Y Le, Lin Y, et al. Anticancer drugs cause release of exosomes with heat shock proteins from human hepatocellular carcinoma cells that elicit effective natural killer cell antitumor responses in vitro. *J Biol Chem*. 2012;287(19):15874-15885. doi:10.1074/JBC.M112.340588

117. Xiao W, Dong W, Zhang C, et al. Effects of the epigenetic drug MS-275 on the release and function of exosome-related immune molecules in hepatocellular carcinoma cells. *Eur J Med Res*. 2013;18(1). doi:10.1186/2047-783X-18-61

118. Ko SF, Yip HK, Zhen YY, et al. Adipose-Derived Mesenchymal Stem Cell Exosomes Suppress Hepatocellular Carcinoma Growth in a Rat Model: Apparent Diffusion Coefficient, Natural Killer T-Cell Responses, and Histopathological Features. *Stem Cells Int*. 2015;2015. doi:10.1155/2015/853506

119. Rao Q, Zuo B, Lu Z, et al. Tumor-derived exosomes elicit tumor suppression in murine hepatocellular carcinoma models and humans in vitro. *Hepatology*. 2016;64(2):456-472. doi:10.1002/HEP.28549

120. Lu Z, Zuo B, Jing R, et al. Dendritic cell-derived exosomes elicit tumor regression in autochthonous hepatocellular carcinoma mouse models. *J Hepatol*. 2017;67(4):739-748. doi:10.1016/J.JHEP.2017.05.019

121. Lötvall J, Hill AF, Hochberg F, et al. Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles. *J Extracell vesicles*. 2014;3(1). doi:10.3402/JEV.V3.26913

122. Witwer KW, Buzás EI, Bemis LT, et al. Standardization of sample collection, isolation and analysis methods in extracellular vesicle research. *J Extracell vesicles*. 2013;2(1). doi:10.3402/JEV.V2I0.20360

123. Vozel D, Uršič B, Krek JL, Štukelj R, Kralj-Iglič V. Applicability of extracellular vesicles in clinical studies. *Eur J Clin Invest*. 2017;47(4):305-313. doi:10.1111/ECI.12733

124. Stahl PD, Raposo G. Exosomes and extracellular vesicles: the path forward. *Essays Biochem*. 2018;62(2):119-124. doi:10.1042/EBC20170088

125. Bastos N, Ruivo CF, da Silva S, Melo SA. Exosomes in cancer: Use them or target them? *Semin Cell Dev Biol*. 2018;78:13-21. doi:10.1016/J.SEMCDB.2017.08.009

PART II

HYPOTHESIS AND OBJECTIVES

Cholangiocarcinoma (CCA) includes a heterogeneous group of biliary cancers with dismal prognosis. These tumors usually have a silent growth and are consequently commonly diagnosed in advanced phases, when the disease is already disseminated, limiting the accessibility to therapeutic options with curative-intent. Current non-invasive diagnostic tools show suboptimal accuracy for CCA identification, particularly at early tumor stages. Consequently, there is an urgent need to develop accurate non-invasive diagnostic tools for the detection of early CCAs, particularly in patients at high-risk. Extracellular vesicles (EVs), small membranous speres found in biofluids which contain biomolecules such as proteins, metabolites, nucleic acids and lipids from the cell of origin, have recently emerged a as a potential source of biomarkers for human disorders. In accordance with this, we have here hypothesized that EVs might constitute a novel source of accurate and non-invasive biomarkers for the prediction, early diagnosis and prognosis estimation in CCA. Therefore, this dissertation aims to describe nucleic acid and protein biomarkers for CCA in serum and urine EVs.

Hence, the following objectives were proposed to be assessed:

I. Isolation and characterization of serum and urine EVs from patients with CCA, PSC and UC as well as from healthy individuals.

II. Determination of the biofluid EV transcriptome and proteome by omic technologies including gene expression array and high-performance liquid chromatography-mass spectrometry, respectively.

III. Identification of serum and/or urine predictive, diagnostic and/or prognostic EV biomarkers for CCA.

IV. Ascertainment of the association between biofluid EV biomolecules with human liver, CCA tissues and tumor-composing cellular subsets in a context of liquid biopsy.

PART III

RNA MOLECULES IN SERUM AND URINE EXTRACELLULAR VESICLES

CHAPTER 3

PATIENTS WITH CHOLANGIOCARCINOMA PRESENT SPECIFIC RNA PROFILES IN SERUM AND URINE EXTRACELLULAR VESICLES MIRRORING THE TUMOR EXPRESSION: NOVEL LIQUID BIOPSY BIOMARKERS FOR DISEASE DIAGNOSIS

Ainhoa Lapitz,[1] Ander Arbelaiz,[1] Colm J O'Rourke,[2] Jose L Lavin,[3] Adelaida La Casta,[1] Cesar Ibarra,[4] Juan P Jimeno,[5] Alvaro Santos-Laso,[1] Laura Izquierdo-Sanchez,[1,6] Marcin Krawczyk,[7,8] Maria J Perugorria,[1,6] Raul Jimenez-Agüero,[1] Alberto Sanchez-Campos,[4] Ioana Riaño,[1] Esperanza Gónzalez,[9] Frank Lammert,[7] Marco Marzioni,[10] Rocio I R Macias,[6,11] Jose J G Marin,[6,11] Tom H Karlsen,[12] Luis Bujanda,[1,6] Juan M Falcón-Pérez,[6,9,13] Jesper B Andersen,[2] Ana M Aransay,[3,6] Pedro M Rodrigues,[1] Jesus M Banales[1,6,13]

[1]Department of Liver and Gastrointestinal Diseases, Biodonostia Health Research Institute, Donostia University Hospital, University of the Basque Country (UPV/EHU), 20014, San Sebastian, Spain.
[2]Biotech Research & Innovation Centre (BRIC), Department of Health and Medical Sciences, 2200, Copenhagen, Denmark.
[3]CIC bioGUNE, Genome Analysis Platform, 48160, Derio, Spain.
[4]Hospital of Cruces, 48903, Bilbao, Spain.
[5]"Complejo Hospitalario de Navarra", 31008, Pamplona, Spain.
[6]Carlos III National Institute of Health, Center for the Study of Liver and Gastrointestinal Diseases (CIBERehd), 28220, Madrid, Spain.
[7]Department of Medicine II, Saarland University Medical Centre, Saarland University, 66421, Homburg, Germany.
[8]Laboratory of Metabolic Liver Diseases, Centre for Preclinical Research, Department of General, Transplant and Liver Surgery, 02-091, Warsaw, Poland.
[9]Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Exosomes Laboratory, 48160, Derio, Spain.
[10]"Università Politecnica delle Marche", Department of Gastroenterology, 60121, Ancona, Italy.
[11]Experimental Hepatology and Drug Targeting (HEVEFARM), Biomedical Research Institute of Salamanca (IBSAL), 37007, Salamanca, Spain.
[12]Norwegian PSC Research Center, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, 0372, Oslo, Spain.
[13]IKERBASQUE, Basque Foundation for Science, 48013, Bilbao, Spain.

Published in *Cells, 2020 Mar 14;9(3):721.*

ABSTRACT

Cholangiocarcinoma (CCA) comprises a group of heterogeneous biliary cancers with dismal prognosis. The etiologies of most CCAs are unknown, but primary sclerosing cholangitis (PSC) is a risk factor. Non-invasive diagnosis of CCA is challenging and accurate biomarkers are lacking. We aimed to characterize the transcriptomic profile of serum and urine extracellular vesicles (EVs) from patients with CCA, PSC, ulcerative colitis (UC), and healthy individuals. Serum and urine EVs were isolated by serial ultracentrifugations and characterized by nanoparticle tracking analysis, transmission electron microscopy, and immunoblotting. EVs transcriptome was determined by Illumina gene expression array [messenger RNAs (mRNA) and non-coding RNAs (ncRNAs)]. Differential RNA profiles were found in serum and urine EVs from patients with CCA compared to control groups (disease and healthy), showing high diagnostic capacity. The comparison of the mRNA profiles of serum or urine EVs from patients with CCA with the transcriptome of tumor tissues from two cohorts of patients, CCA cells in vitro, and CCA cells-derived EVs, identified 105 and 39 commonly-altered transcripts, respectively. Gene ontology analysis indicated that most commonly-altered mRNAs participate in carcinogenic steps. Overall, patients with CCA present specific RNA profiles in EVs mirroring the tumor, and constituting novel promising liquid biopsy biomarkers.

1. INTRODUCTION

Cholangiocarcinomas (CCAs) are heterogeneous biliary malignancies characterized by dismal prognosis. The incidence and mortality rates of these cancers are rapidly increasing globally, currently accounting for ~15% of all primary liver cancers and ~3% of gastrointestinal malignancies.[1–3] According to their anatomical localization, CCAs are classified into intrahepatic (iCCA), perihilar (pCCA), or distal (dCCA). The etiology of most CCAs is unknown. However, several risk factors have been described, including the presence of primary sclerosing cholangitis (PSC: 5–15% develops CCA), a chronic cholestatic liver disease that is associated with autoimmune phenomena against the intra- and extrahepatic bile ducts.[1,4,5] Importantly, 70–80% of patients with PSC concomitantly present inflammatory bowel disease (IBD), mainly ulcerative colitis (UC), which is thought to precede the development of the liver disease.[4]

CCAs are generally asymptomatic in early stages, being therefore commonly diagnosed in advanced phases when the disease is disseminated. Late diagnosis combined with the chemoresistant nature of these tumors highly compromise the current therapeutic options, mainly based on surgery, significantly impacting on patient's welfare and outcome.[1,2,6] The diagnosis of CCA is usually conducted by combining clinical, biochemical, radiological, and histological information. Imaging techniques usually rely on computed tomography (CT), magnetic resonance imaging (MRI), magnetic resonance cholangiopancreatography (MRCP), positron emission tomography (PET), percutaneous transhepatic cholangiography (PTC), endoscopic retrograde cholangiopancreatography (ERCP), or endoscopic ultrasound depending on the tumor location.[1,5,7] However, imaging proceedings have important limitations, as they are not accurate enough to determine the malignity of the tumor masses, particularly in early stages, as well as to differentiate between the main primary liver cancers, *i.e.*, iCCA, hepatocellular carcinoma (HCC) or HCC-CCA mixed tumors, which is fundamental to provide the appropriate standards of care. On the other hand, MRCP and histological analysis (biopsy or brushing cytology) comprise the major diagnostic tools for PSC.[4,8,9] In addition, the measurement of non-specific serum tumor biomarkers [*i.e.*, carbohydrate antigen 19-9 (CA19-9) and carcinoembryonic antigen (CEA)] is commonly conducted in order to help in the diagnosis of CCA, but their low sensitivity (particularly in early stages of the disease) and specificity (also elevated in some PSC patients without cancer), raise important concerns regarding their clinical utility.[7,10] Therefore, tumor biopsy is currently mandatory to confirm the diagnosis and staging of CCA, guiding the clinical management of these patients.[11] Based on all of these diagnostic concerns, there is an urgent need to determine new accurate, non-invasive biomarkers for the early diagnosis of CCA, particularly in patients at risk.

In the last decade, extracellular vesicles (EVs) have been envisioned as promising tools in the quest for tumor biomarkers and as important mediators of disease pathogenesis.[12] EVs constitute a heterogeneous population of lipid bilayered spheres (30 nm – 2 μm in diameter) containing diverse biomolecules (*e.g.*, proteins, nucleic acids, lipids, and metabolites), which are released by cells and found in all biofluids (*e.g.*, blood and urine).[13,14] Taking into account their biogenesis, EVs may be classified as exosomes, microvesicles (MV), and apoptotic bodies. These small vesicles participate in cell-to-cell communications, modulating signalling pathways in pathobiology.[13–17] We previously reported a differential proteomic profile of serum EVs from patients with CCA, HCC, or PSC, as well as from healthy individuals, identifying accurate candidate biomarkers for the differential diagnosis of these diseases.[18] Considering that tumor cells can also release RNAs encapsulated within EVs, and that their profiles can mimic the cellular state/alterations, an extensive characterization of the RNA content from serum and urine EVs from patients with CCA (and control conditions) might provide new diagnostic biomarkers as well as therapeutic targets.

In this study, we aimed to characterize the RNA profile of serum and urine EVs from patients with CCA, PSC, or UC, as well as healthy individuals, and identify candidate diagnostic biomarkers mirroring their tumor cell expression within the liquid biopsy concept. For this purpose, the expression of selected candidates was evaluated in human CCA tumor and surrounding healthy tissues from two independent cohorts of patients [The Cancer Genome Atlas (TCGA) and Copenhagen], as well as in cell cultures (CCA *vs*. normal) and EVs released by normal or tumor human cholangiocytes in vitro.

2. MATERIALS AND METHODS

2.1 Patients

Serum and urine samples from patients with CCA (n = 12 and 23, respectively), PSC (n = 6 and 5, respectively), UC (n = 8 and 12, respectively), and healthy individuals (n = 9 and 5, respectively) were obtained from Donostia University Hospital (San Sebastian, Spain), Cruces University Hospital (Bilbao, Spain), and “Complejo Hospitalario de Navarra” (Pamplona, Spain). The Ethical Committees for Clinical Research from each participating institution approved all the research protocols and all patients accepted to participate in the study and signed the written consents to allow the use of their samples for biomedical research. Clinical characteristics of patients and tumors are summarized in **Table 3.1**. The diagnosis of PSC was based on the European Association for the Study

of the Liver (EASL) guidelines by demonstrating the presence of bile duct alterations (strictures or irregularities in intrahepatic and extrahepatic bile ducts) using MRCP after excluding secondary causes of cholangitis.[19] The diagnosis of UC was performed by combining endoscopic and histological studies, mainly colonoscopy in parallel with hematoxylin and eosin (H&E) staining, after excluding other potential diseases. Finally, CCA diagnosis was confirmed by histological analysis of tumor samples and/or through the combination of clinical, biochemical and radiological approaches. Tumor stage was determined based on the 7th edition of the American Joint Committee on Cancer (AJCC) classification.

RNA-seq data from the TCGA cohort (36 CCAs and 9 surrounding liver samples),[20] downloaded as level 3 data through FireBrowse portal [BROAD Institute of MIT & Harvard, MA, USA (source: https://gdac.broadinstitute.org/)], and whole transcriptome profiling [Human Transcriptome (HT) BeadChips (Illumina Inc., San Diego, CA, USA)] of the “Copenhagen cohort” including 217 CCA surgical specimens (153 iCCA, 43 pCCA, 15 dCCA, 6 unknown location), 143 normal surrounding liver samples, and 9 normal intrahepatic bile ducts (GSE26566) were used to evaluate the expression of serum and urine biomarkers in tumor tissue.[21,22]

Table 3.1. Clinical information from study individuals

		Serum				Urine			
		Healthy	UC	PSC	CCA	Healthy	UC	PSC	CCA
Patients (n)		9	8	6	12	5	12	5	23
Age, mean [range] (years)		**41.4** [26-72]	**54.3** [33-82]	**45.5** [16-75]	**67.4** [59-79]	**66.2** [53-72]	**58.8** [42-82]	**49.6** [26-75]	**67.8** [50-82]
Gender, M / F (n/n)		4 / 5	5 / 3	6	9 / 3	3 / 2	6 / 6	4 / 1	16 / 7
ALT, mean [range] (IU/L)		NA	**16.50** [9-29]	**44.33** [12-128]	**56.67** [11-200]	NA	**15.83** [9-30]	**36.60** [18-62]	**115.96** [11-672]
AST, mean [range] (IU/L)		NA	**16.25** [10-20]	**36.83** [16-83]	**46.09** [19-104]	NA	**17.67** [10-28]	**35.80** [20-56]	**85.00** [19-379]
GGT, mean [range] (IU/L)		NA	**24.88** [13-38]	**122.33** [17-361]	**750.33** [87-3011]	NA	**22.25** [14-38]	**119.40** [35-241]	**781.87** [87-3011]
Bilirubin, mean [range] (mg/dL)		**0.85** [0.4-1.3]	**0.68** [0.3-1.1]	**1.10** [0.5-2.4]	**2.65** [0.3-16.7]	**0.68** [0.4-1.3]	**0.63** [0.3-1.1]	**1.00** [0.3-2.4]	**6.00** [0.3-31.3]
ALP, mean [range] (mg/dL)		**57.50** [52-63]	**60.38** [42-95]	**237.50** [54-634]	**320.33** [68-699]	**70.00** [52-95]	**67.67** [42-95]	**301.8** [54-634]	**406.91** [68-1956]
CA19-9, mean [range] (U/mL)		**12.95** [5.6-38.5]	NA	**16.33** [5-35]	**5346.10** [51.5-22039]	**13.18** [4.9-38.5]	NA	**13.33** [0-35]	**5967.60** [51.5-23244]
CCA type	iCCA	-	-	-	4	-	-	-	5
	pCCA	-	-	-	6	-	-	-	9
	dCCA	-	-	-	2	-	-	-	9
Tumor stage	I	-	-	-	1	-	-	-	1
	II	-	-	-	3	-	-	-	7
	III	-	-	-	4	-	-	-	4
	IV	-	-	-	3	-	-	-	8
	Unknown	-	-	-	1	-	-	-	3
Tumor differentiation grade	Well	-	-	-	2	-	-	-	6
	Moderate	-	-	-	1	-	-	-	4
	Poor	-	-	-	0	-	-	-	0
	Unknown	-	-	-	9	-	-	-	13
IBD	No IBD	0	0	1	12	0	0	1	23
	UC	0	8	3	0	0	12	2	0
	Crohn	0	0	1	0	0	0	2	0
	Unspecified	0	0	1	0	0	0	0	0

2.2 Cell cultures

Normal human cholangiocytes (NHCs) were isolated from normal liver tissue and characterized as previously described.[23–25] Furthermore, two commercial human CCA cell lines (EGI1 and TFK1, Leibniz Institute DSMZ-German Collection of Microorganism and Cell Cultures, Germany) were used. NHC and EGI1 cells were cultured in fully-supplemented DMEM/F-12 medium, as previously described,[23–25] while TFK1 cells were cultured in DMEM/F-12 supplemented with 10% fetal bovine serum (FBS; Gibco, Thermo

Fisher Scientific, Waltham, MA, USA) and 1% penicillin/streptomycin (P/S; Gibco). Cells were seeded in 150 mm collagen-coated tissue culture dishes (4 × 106 cells) with each respective cell culture medium and left for plate attachment overnight. Afterwards, cells were washed with phosphate-buffered saline (PBS) and incubated with “EV recollection media” (DMEM/F-12+Glutamax supplemented with 1% P/S, and without serum). After 48 h, cells were harvested for RNA isolation and cell culture media was collected and stored at −80 °C for subsequent EVs isolation. Cells were grown at 37 °C in a humidified chamber of 5% CO2. During all the experiments, mycoplasm test was performed by conventional PCR and cells were tested as mycoplasm negative.

2.3 Isolation of EVs from serum, urine, and cell cultures

Serum, urine, and cell-derived EVs were isolated as previously described.[18] Briefly, 1 mL of serum, 50 mL of urine, or 300 mL of cell culture media (frozen at −80 °C) were thawed at room temperature and further processed through serial differential ultracentrifugation steps at 4 °C. First, in order to remove cell debris, serum, urine and cell culture media were centrifuged at 10,000× g for 30 min and subsequently ultracentrifuged at 100,000× g for 75 min, to pellet the EVs, which were then washed with PBS and pelleted again after ultracentrifugation at 100,000× g for 75 min. Finally, the pelleted EV fraction was resuspended in 20 µL of PBS and then stored at −80 °C for further analysis.

2.4 Transmission Electron Microscopy (TEM)

For the characterization of EVs, the isolated fraction of EVs was stained negatively and analyzed by TEM. EV samples were directly adsorbed onto glow-discharged (60 seg low discharging using a PELCO easy-glow device) carbon-coated copper grid (300 mesh). Afterwards, grids were fixed with 2% paraformaldehyde (PFA) in phosphate buffer (PB 0.2M pH 7.4) for 20 min and washed with distilled water. Then, the contrast staining was made by incubating the grids with 4% uranyl acetate (UA) at 4 °C for 15 min. TEM images were obtained by using TECNAI G2 20 C-TWIN high-resolution transmission electron microscope, at an acceleration voltage of 200 kV.

2.5 Immunoblotting

Protein levels of both EV and endoplasmic reticulum markers (*i.e.*, CD63 and CD81 *vs*. GRP78, respectively) were evaluated in serum and urine EVs and in whole-cell extracts (WCEs) by immunoblotting. Total protein concentration was calculated with the Micro BCA protein assay kit (Thermo Fisher Scientific,), following the manufacturer's instructions. Loading buffer [50 mM Tris-HCl, 2% SDS, 10% glycerol and 0.1% bromophenol blue, without β-mercaptoethanol or dithiothreitol (DTT)] was added to protein samples, followed by heat denaturation at 95 °C for 5 min. Then, 10 and 4 µg of total protein from serum and urine EVs, respectively, were separated by 12.5% sodium dodecyl sulfate-polyacrilamide gel electrophoresis (SDS-PAGE) and electro-transferred onto a nitrocellulose membrane (GE Healthcare, Chicago, IL, USA) and blocked with 5% skim milk powder/tris-buffered saline (TBS)-0.1% tween (TBS-Tween) for 1 h. Afterwards, membranes were probed overnight at 4°C with the appropriate primary antibodies [anti-CD81 (BD Biosciences), anti-CD63 (DSHB), and anti-GRP78 (BD Biosciences, San Jose, CA, USA)] at 1:500 dilution in blocking solution and, after three washes with TBS-Tween (5 min each), horseradish peroxidase-conjugated secondary antibody (anti-mouse; Cell Signaling, Danvers, MA, USA) at a dilution of 1:5000 (in milk blocking solution) were incubated for 1 h at room temperature. Membranes were developed for protein detection using ECL plus (Thermo Fisher Scientific), with the iBright FL1500 Western Blot Imaging System (Thermo Fisher Scientific).

2.6 EV Size and Concentration

Size distribution and concentration of EVs were evaluated by nanoparticle tracking analysis (NTA) using a NanoSight LM10 System (Malvern, UK) further equipped with fast video capture and a particle-tracking software. NTA post-acquisition settings were kept constant for all samples. Each video was analyzed for obtaining the mean and mode vesicle size as well as particle concentration.

2.7 Total RNA Isolation

After EVs isolation, total RNA was extracted using the miRCURY™ RNA Isolation Kit (Qiagen, Hilden, Germany) following manufacturer's specifications. Afterwards, total RNA was resuspended in 20 µL of distilled H2O and later used for transcriptomic analysis. Regarding cell samples, total RNA was extracted using the TRIzol® reagent

according to the manufacturer's instructions (Life Technologies Corp., Carlsbad, CA, USA).

2.8 Illumina Gene Expression Array

Illumina HumanHT-12 WG-DASL V4.0 R2 expression beadchips were used to characterize gene expression [messenger RNAs (mRNAs) and non-coding RNAs (ncRNAs)]. The quality of RNA samples was measured using a RNA Pico Chip Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). 200 ng of RNA samples were used for the array. The cDNA synthesis, prequalification, amplification, labelling and hybridization of the samples were performed following the WG-DASL HT Assay Lab protocol (Illumina Inc.). The amplified cDNAs were hybridized to the diverse gene-probes of the array and gene expression levels were detected by a HiScan scanner (Illumina Inc.). Raw data were extracted with GenomeStudio analysis software (Illumina Inc.), in the form of GenomeStudio's Final Report. Raw expression data were background-corrected, log2-transformed and quantile-normalized using the lumi R package (Bioconductor repository, Chicago, IL, USA).[26] To perform the Venn diagrams, all the transcripts identified in at least one sample with a "detection p-value" < 0.01 were selected. Afterwards, in the comparisons between groups, transcripts that were significantly identified in at least 20% of the samples (with a bilateral p-value < 0.05; independent samples two-tailed t-test, not assuming equal variances) were considered for subsequent analysis.

2.9 Functional Enrichment Analysis

Functional analysis of candidate liquid biopsy RNA biomarkers was determined by gene ontology (GO) enrichment of biological processes, molecular pathways and functions, by using the Functional Enrichment analysis tool (FunRich) version 3.1.3 (Funrich Industrial Co. Ltd, Hong Kong).[27,28]

2.10 Statistical Analysis

Statistical analysis was performed using GraphPad Prism version 6.0 (GraphPad Software, San Diego, CA, USA). Data are shown as boxes and whiskers (min to max). When comparing two groups, non-parametric Mann-Whitney or parametric t-Student

tests were conducted. For comparisons between more than two groups, non-parametric Kruskal-Wallis test followed by a posteriori Dunns test of the parametric one-way analysis of variance (ANOVA) test followed by a posteriori Tukey's post hoc test were used. In order to calculate the diagnostic values of serum and urine RNA biomarkers, allowing to discriminate between patients with CCA, PSC and UC, and healthy individuals, area under the receiver operating characteristic curve (AUC) values were determined using the SPSS 20.0 software (IBM, Ehningen, Germany), followed by the calculation of sensitivity (SEN) and specificity (SPE) values, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR), negative likelihood ratio (NLR), and accuracy index (AI). Differences were considered significant when $p < 0.05$.

3. RESULTS

3.1 Characterization of Serum and Urine EVs from Patients with CCA, PSC, or UC, and Healthy Individuals

After isolation, serum and urine EVs were characterized by TEM, immunoblotting and NTA. In resemblance with our previous findings using the same isolation protocol,[18] TEM images showed a typical rounded morphology in the isolated vesicles from both serum and urine (~100–200 nm), corresponding to exosomes and/or small microvesicles (**Figure 3.1A**). By immunoblotting, the EV protein markers CD63 and CD81 were highly enriched in the isolated EV fraction, when compared to total serum or NHC whole-cell extracts (WCE), while the endoplasmic reticulum marker 78 kDa glucose-regulated protein (GRP78) was completely absent in isolated serum and urine EVs but only found expressed in WCE from NHCs (**Figure 3.1B**), substantiating a proper isolation and a high purity of the obtained EVs. Regarding the size of EVs, NTA revealed no significant differences in the size of serum and urine EVs among groups, presenting an average size of ~180 nm, in resemblance with serum and urine EV concentration, which was found similar in the study population (**Figure 3.1C**).

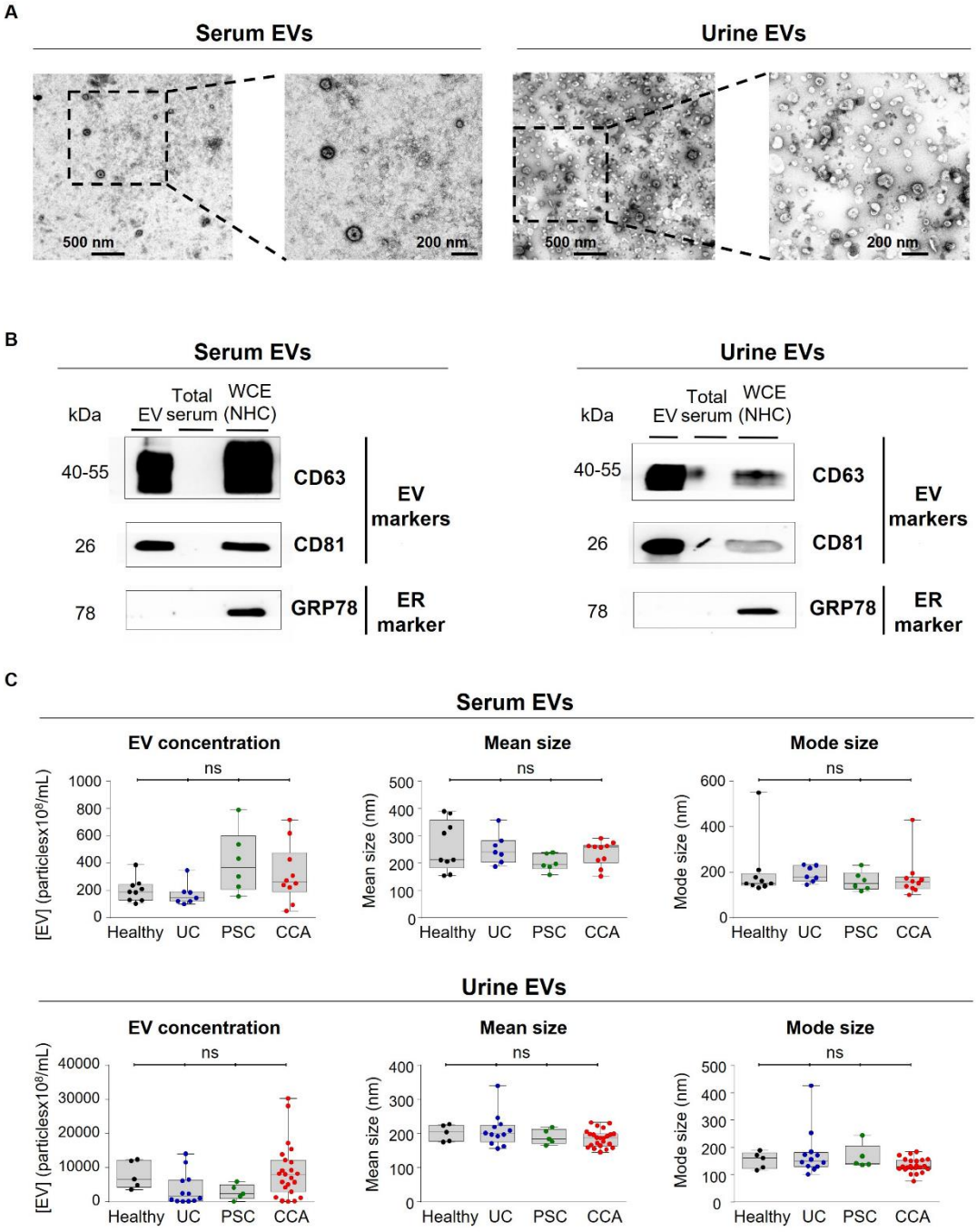


Figure 3.1. Characterization of serum and urine EVs from patients with CCA, PSC, UC, and healthy controls. In order to validate the protocol for EVs isolation, we used blood serum and urine from healthy individuals. (**A**) TEM images of blood serum (left) and urine (right) EVs from healthy individuals showcasing the typical round shape (~150 nm) and morphology. (**B**) Representative immunoblots of the EV markers CD63 and CD81 (positive controls) and GRP78 (negative control) from EVs isolated from serum (left) and urine (right) of healthy individuals that indicate an enrichment of EV markers and a complete absence of the endoplasmic reticulum (ER) marker GRP78, compared to total serum and whole cell extracts (WCEs) of normal human cholangiocytes (NHC). (**C**) Nanoparticle tracking analysis (NTA) of serum (up) and urine (down) EVs revealing no differences in EV concentration between CCA, PSC, UC, and healthy individuals and a similar EV mode (~180 nm).

3.2 Differential RNA Profiles of Serum EVs from CCA, PSC, UC, and Healthy Individuals

The transcriptomic profiles of serum and urine EVs isolated from patients with CCA, PSC, UC, and healthy controls were determined by RNA microarray-based transcriptomics (Illumina Inc.). Transcriptomic data are available in GSE144521. Considering all the transcripts that were identified in at least one sample included in any of the study groups (detection p-value < 0.01), a total of 25,084 transcripts were identified in serum EVs. Among them, 10,104 transcripts were identified in serum EVs isolated from healthy individuals, in parallel with the identification of 11,124, 4204, and 24,264 transcripts in serum EVs from patients with UC, PSC, and CCA, respectively, with 1617 of the identified transcripts being shared among all groups (**Figure 3.2A**). In all the study groups, the great majority of the identified transcripts were mRNAs (9516, 10,526, 3949, and 23,029 transcripts found in healthy individuals and patients with UC, PSC, or CCA, respectively), followed by non-coding RNAs such as non-coding RNAs (mainly including pseudogenes, long non-coding RNAs (lncRNAs), among others), microRNAs (miRNAs or miRs), and small nucleolar RNAs (snoRNAs) (**Figure 3.2B**; **Table 3.2**). Other types of RNAs, including small nuclear, miscellaneous, guide, small cytoplasmic, antisense, RNase MRP, ribosomal, and telomerase RNAs were also detected.

Next, the transcriptome of serum EVs from the four study groups was determined and compared. Specifically, 1932 transcripts were differentially identified in CCA *vs*. healthy individuals, 2888 in CCA *vs*. PSC, and 2807 in CCA *vs*. “PSC, UC, and healthy individuals” combined as one unique control (disease and healthy) group (**Figure 3.3**). Meanwhile, 866 transcripts were differentially identified in serum EVs from patients with PSC compared with a group comprised of patients with UC and healthy individuals (**Figure 3.4**).

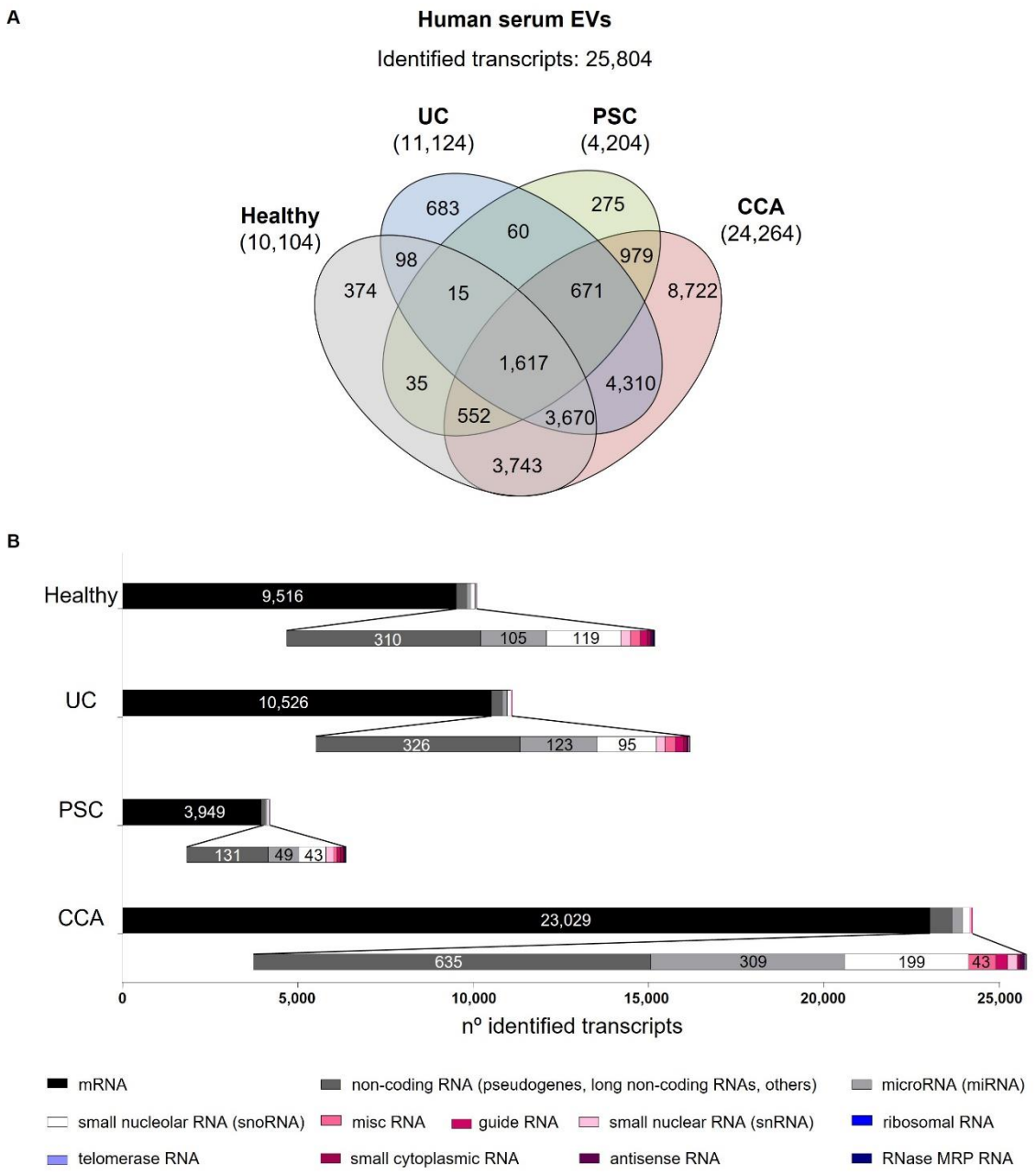
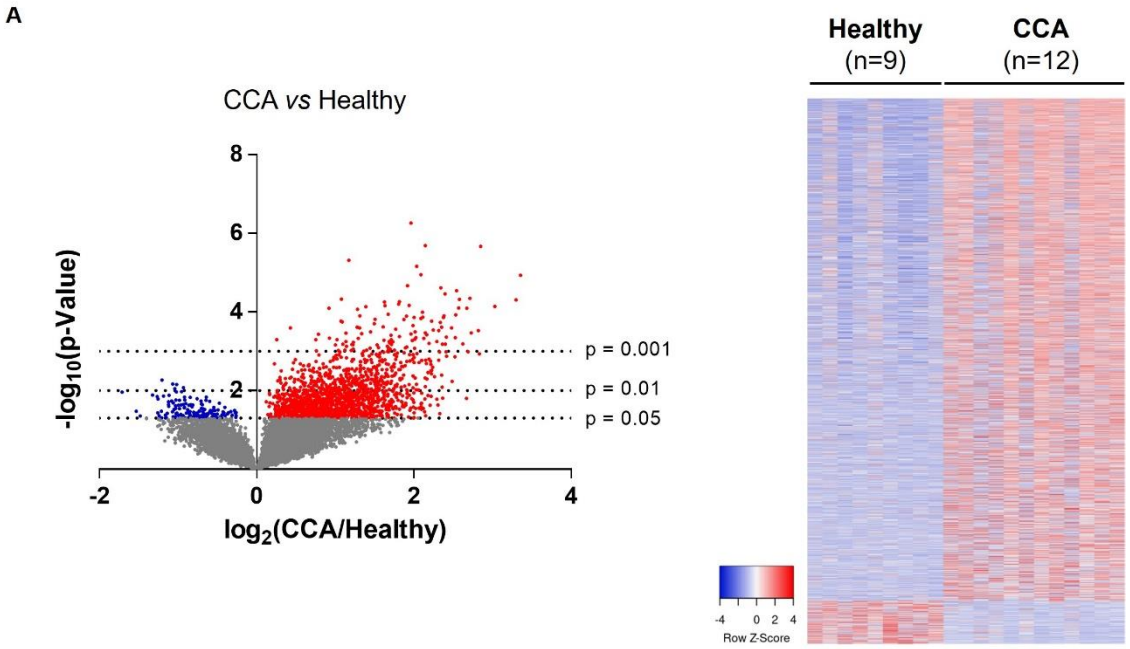


Figure 3.2. Comparative transcriptomic analysis of serum EVs from patients with CCA, PSC, or UC, and healthy individuals. (**A**) Venn diagrams showing the number of transcripts identified per group. (**B**) Number of transcripts identified within each study group, subclassified according to their type [messenger RNA (mRNA), non-coding RNA (including mostly pseudogenes and long non-coding RNAs, among others), miscelaneous RNA (miscRNA), guide RNA, microRNA (miRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), ribosomal RNA, telomerase RNA, small cytoplasmic RNA, antisense RNA and RNase MRP RNA]. In all groups, mRNAs constitute the most abundantly identified RNAs, followed by non-coding RNAs (pseudogenes, lncRNAs, and others), miRNAs, and snoRNAs.

Table 3.2. Number of transcripts identified in each experimental group

	Array transcripts	Serum				Urine			
		Healthy	UC	PSC	CCA	Healthy	UC	PSC	CCA
mRNA	27,580	9,516	10,526	3,949	23,029	10,949	22,801	21,145	24,817
non-coding RNA	726	310	326	131	635	202	581	542	637
microRNA	518	105	123	49	309	57	273	224	319
small nucleolar RNA	263	119	95	43	199	70	175	162	201
misc RNA	48	16	16	6	43	14	39	39	42
guide RNA	23	11	13	5	19	5	20	16	20
small nuclear RNA	16	15	14	12	15	14	15	15	15
antisense RNA	10	4	4	2	7	4	8	8	7
small cytoplasmic RNA	5	5	5	5	5	5	5	5	5
RNase MRP RNA	1	1	1	1	1	1	1	1	1
ribosomal RNA	1	1	1	1	1	1	1	1	1
telomerase RNA	1	1	0	0	1	1	1	1	1
		Total: 25,804				**Total: 27,319**			

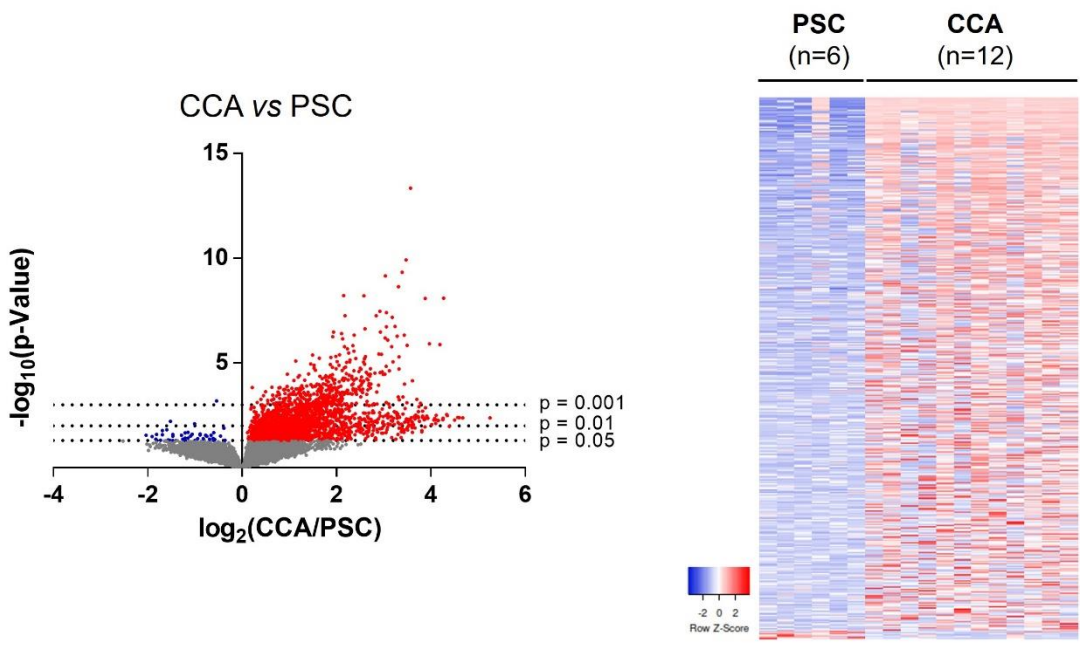


Transcript		CCA (n=12) *vs* Healthy (n=9)								
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
RFFL	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
OR4F3	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
FAM107B	mRNA	**0.991** (0.961, 1.000)	<0.001	91.7	100.0	100.0	90.0	>100	0.08	95.3
ZNF266	mRNA	**0.991** (0.961, 1.000)	<0.001	91.7	100.0	100.0	90.0	>100	0.08	95.3
CADM1	mRNA	**0.981** (0.935, 1.000)	<0.001	91.7	100.0	100.0	90.0	>100	0.08	95.3
CCNB1IP1	mRNA	**0.981** (0.935, 1.000)	<0.001	91.7	100.0	100.0	90.0	>100	0.08	95.3
POTEH	mRNA	**0.981** (0.935, 1.000)	<0.001	91.7	100.0	100.0	90.0	>100	0.08	95.3
RPL26L1	mRNA	**0.981** (0.934, 1.000)	<0.001	100.0	88.9	92.3	100.0	9.00	0.00	95.2
OR5B3	mRNA	**0.972** (0.914, 1.000)	<0.001	83.3	100.0	100.0	81.8	>100	0.17	90.5
GAD1	mRNA	**0.972** (0.909, 1.000)	<0.001	91.7	100.0	100.0	90.0	>100	0.08	95.2

Transcript		CCA (n=12) *vs* Healthy (n=9)								
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
PMS2L4	pseudogene	**0.991** (0.961, 1.000)	<0.001	91.7	100.0	100.0	90.0	>100	0.08	95.2
miR-604	miRNA	**0.944** (0.851, 1.000)	<0.001	83.3	100.0	100.0	81.8	>100	0.17	90.5
SNORA58	snoRNA	**0.926** (0.817, 1.000)	<0.001	75.0	100.0	100.0	75.0	>100	0.25	85.7
LOC100131726	lncRNA	**0.907** (0.783, 1.000)	<0.01	75.0	100.0	100.0	75.0	>100	0.25	85.7
miR-1224	miRNA	**0.907** (0.782, 1.000)	<0.01	91.7	77.8	84.6	87.5	4.12	0.11	85.7

Figure 3.3. Differential transcriptomic profile of serum EVs and diagnostic capacity. Volcano plot (−log10(p-value) and log2(fold-change); up left), heatmap of the differentially expressed transcripts (up right) and diagrams with the diagnostic capacity with the highest AUC values of the 10 selected mRNAs and 5 selected non-coding RNAs in serum EVs from (**A**) CCA *vs*. Healthy individuals; (**B**) CCA *vs*. PSC; (**C**) CCA *vs*. (PSC + UC + Healthy individuals). Abbreviations: AI, accuracy index; AUC, area under the receiver operating characteristic curve; CI, confidence interval; lncRNA, long non-coding RNA; miRNA, microRNA; NLR, negative likelihood ratio; NPV, negative predictive value; PLR, positive likelihood ratio; PPV, positive predictive value; SEN, sensitivity; snoRNA, small nucleolar RNA; SPE, specificity.

B

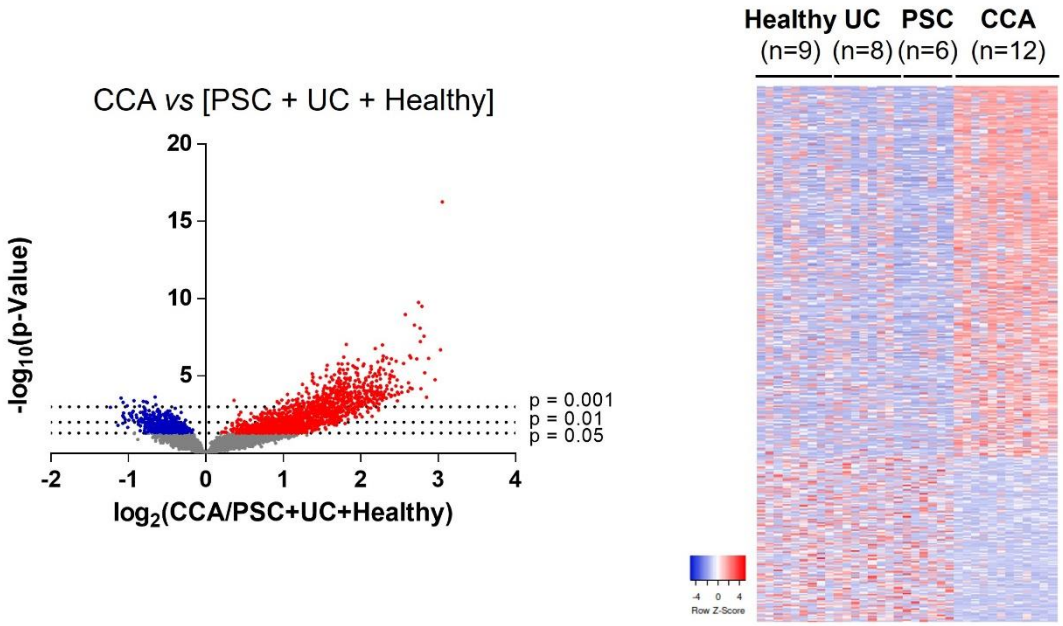


Transcript		CCA (n=12) *vs* PSC (n=6)								
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
PON1	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
ATF4	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
PHGDH	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
CMIP	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
CCNI	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
IRX3	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
DARS	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
GRM3	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
RFFL	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
ZNF148	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0

Transcript		CCA (n=12) *vs* PSC (n=6)								
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
MALAT1	lncRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
SNORA11B	snoRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
LOC100190986	lncRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
GTF2IRD2P	pseudogene	**0.986** (0.943, 1.000)	<0.001	91.7	100.0	100.0	85.7	>100	0.08	94.4
TFAMP1	pseudogene	**0.972** (0.904, 1.000)	<0.001	91.7	100.0	100.0	85.7	>100	0.08	94.4

Figure 3.3. (continued)

C

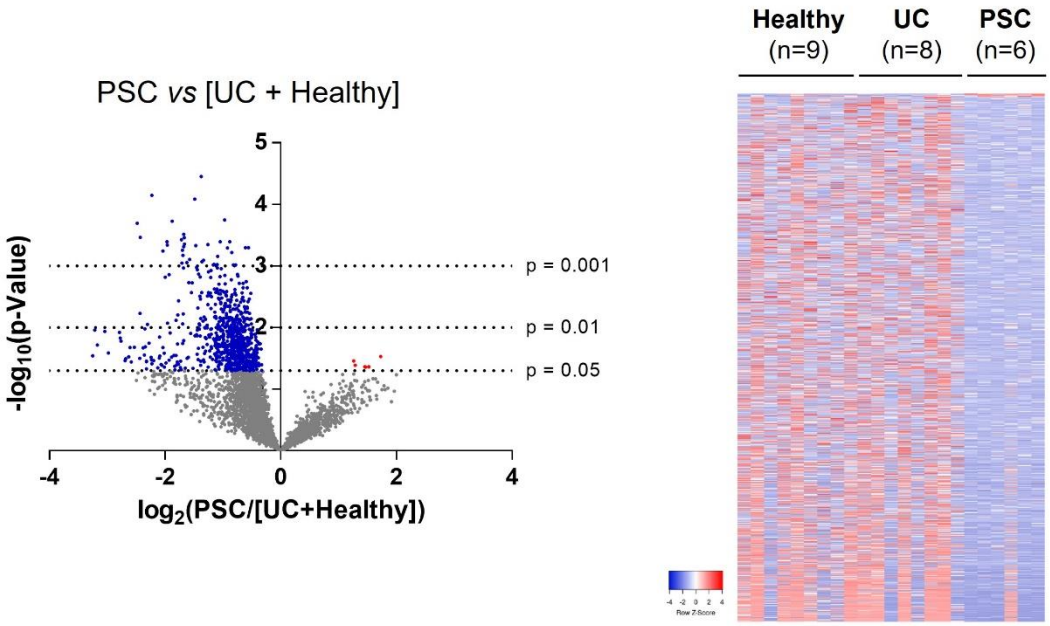


Transcript		CCA (n=12) *vs* [PSC + UC + Healthy] (n=23)								
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
RFFL	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
ZNF266	mRNA	**0.976** (0.936, 1.000)	<0.001	91.7	91.3	84.6	95.5	10.54	0.09	91.4
OR4F3	mRNA	**0.960** (0.901, 1.000)	<0.001	100.0	87.0	80.0	100.0	7.67	0.00	91.4
CMIP	mRNA	**0.957** (0.884, 1.000)	<0.001	100.0	91.3	85.7	100.0	11.5	0.00	94.3
FLJ40453	mRNA	**0.953** (0.882, 1.000)	<0.001	91.7	91.3	84.6	95.5	10.54	0.09	91.4
PON1	mRNA	**0.949** (0.882, 1.000)	<0.001	91.7	91.3	84.6	95.5	10.54	0.09	91.4
FILIP1L	mRNA	**0.946** (0.875, 1.000)	<0.001	100.0	87.0	80.0	100.0	7.67	0.00	91.4
C19orf31	mRNA	**0.944** (0.872, 1.000)	<0.001	83.3	91.3	83.3	91.3	9.58	0.18	88.6
FHL2	mRNA	**0.942** (0.859, 1.000)	<0.001	100.0	87.0	80.0	100.0	7.67	0.00	91.4
PTP4A2	mRNA	**0.938** (0.863, 1.000)	<0.001	75.0	95.7	90.0	88.0	17.25	0.26	88.6

Transcript		CCA (n=12) *vs* [PSC + UC + Healthy] (n=6)								
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
miR-551B	miRNA	**0.909** (0.815, 1.000)	<0.001	83.3	87.0	76.9	90.9	6.39	0.19	85.7
PMS2L4	pseudogene	**0.880** (0.758, 1.000)	<0.001	91.7	87.0	78.6	95.2	7.03	0.10	88.6
LOC643955	pseudogene	**0.873** (0.754, 0.992)	<0.001	83.3	87.0	76.9	90.9	6.39	0.19	85.7
LOC100134868	lncRNA	**0.864** (0.747, 0.982)	<0.001	100.0	65.2	60.0	100.0	2.88	0.00	77.1
PTTG3P	pseudogene	**0.859** (0.712, 1.000)	<0.001	83.3	91.3	83.3	91.3	9.58	0.18	88.6

Figure 3.3. (continued)

A



Transcript		PSC (n=6) *vs* [UC + Healthy] (n=17)								
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
DPY30	mRNA	**0.931** (0.829, 1.000)	<0.01	100.0	82.4	66.7	100.0	5.67	0.00	87.0
KCNJ6	mRNA	**0.931** (0.824, 1.000)	<0.01	100.0	88.2	75.0	100.0	8.50	0.00	91.3
FAM54A	mRNA	**0.931** (0.824, 1.000)	<0.01	100.0	82.4	66.7	100.0	5.67	0.00	87.0
HAMP	mRNA	**0.931** (0.812, 1.000)	<0.01	83.3	94.1	83.3	94.1	14.17	0.18	91.3
PON1	mRNA	**0.922** (0.811, 1.000)	<0.01	100.0	76.5	60.0	100.0	4.25	0.00	82.6
GPR116	mRNA	**0.922** (0.810, 1.000)	<0.01	100.0	76.5	60.0	100.0	4.25	0.00	82.6
MMP21	mRNA	**0.922** (0.810, 1.000)	<0.01	100.0	76.5	60.0	100.0	4.25	0.00	82.6
DCTN1	mRNA	**0.922** (0.807, 1.000)	<0.01	100.0	88.2	75.0	100.0	8.50	0.00	91.3
SLC22A1	mRNA	**0.912** (0.788, 1.000)	<0.01	100.0	88.2	75.0	100.0	8.50	0.00	91.3
PLEKHA6	mRNA	**0.902** (0.776, 1.000)	<0.01	100.0	82.4	66.7	100.0	5.67	0.00	87.0

Transcript		PSC (n=6) *vs* [UC + Healthy] (n=17)								
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
SNORA11B	snoRNA	**0.902** (0.775, 1.000)	<0.01	100.0	76.5	60.0	100.0	4.25	0.00	82.6
MIR632	miRNA	**0.873** (0.717, 1.000)	<0.01	83.3	76.5	55.6	92.9	3.54	0.22	78.3
MIR222	miRNA	**0.873** (0.673, 1.000)	<0.01	83.3	88.2	71.4	93.8	7.08	0.19	87.0
SNORA29	snoRNA	**0.833** (0.654, 1.000)	<0.05	100.0	64.7	50.0	100.0	2.83	0.00	73.9
LOC729603	pseudogene	**0.833** (0.650, 1.000)	<0.05	83.3	76.5	55.6	92.9	3.54	0.22	78.3

Figure 3.4. Differential transcriptomic profile and diagnostic efficacy of serum EV transcripts. (**A**) Volcano plot [-log10(p-value) and log2(fold-change); up left], heatmap of the differentially expressed transcripts (up right) and diagrams with the diagnostic capacity with the highest AUC values of the 10 selected mRNAs and 5 selected non-coding RNAs in serum EVs from PSC *vs* [UC + Healthy individuals]. (**B**) Potential serum EV biomarkers for PSC. Abbreviations: AI, accuracy index; AUC, area under the receiver operating characteristic curve; CI, confidence interval; lncRNA, long non-coding RNA; miRNA, microRNA; NLR, negative likelihood ratio; NPV, negative predictive value; PLR, positive likelihood ratio; PPV, positive predictive value; SEN, sensitivity; snoRNA, small nucleolar RNA; SPE, specificity.

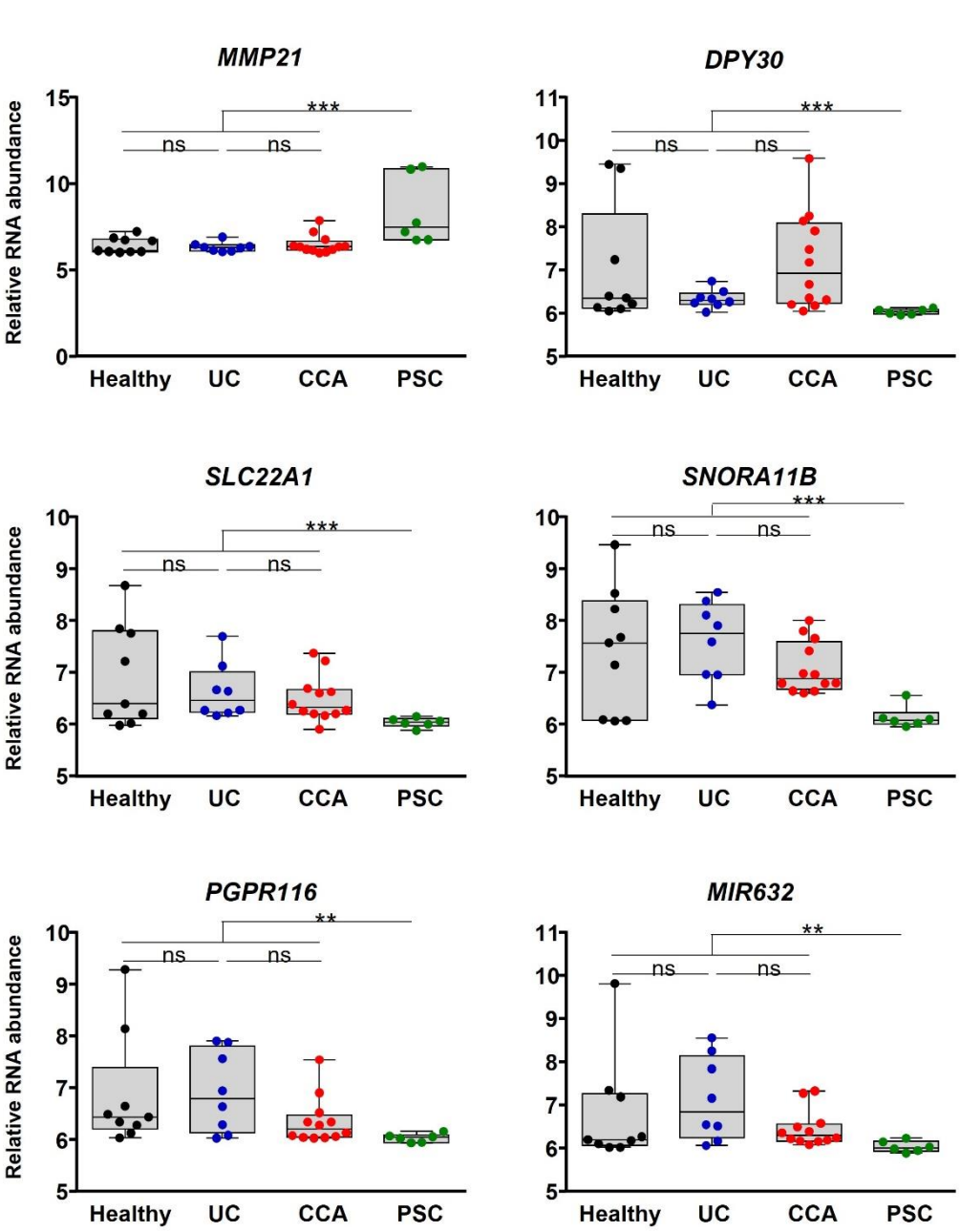


Figure 3.4. (Continued)

The analysis of candidate RNA biomarkers in serum EVs from CCA *vs*. healthy individuals pointed out ring finger and FYVE like domain containing E3 ubiquitin protein ligase (RFFL), olfactory receptor family 4 subfamily F member 3 (OR4F3), and the family with sequence similarity 107 member B (FAM107B) as the mRNAs with the highest diagnostic capacity, presenting AUC values of 1.00, 1.00, and 0.991, respectively, along with the non-coding RNAs PMS1 homolog 2 mismatch repair system component pseudogene 4 (PMS2L4), miR-604, and SNORA58 (AUC: 0.991, 0.944, and 0.926, respectively) (**Figure 3.3A**). Since PSC is a well-known risk factor that increases the odds of developing CCA, the transcriptomic profiles of serum EVs from patients with

CCA *vs*. PSC were also compared. In particular, the mRNA transcripts paraoxonase 1 (PON1), activating transcription factor 4 (ATF4), and phosphoglycerate dehydrogenase (PHGDH) stood out as the best candidate biomarkers for the differential diagnosis of CCA and PSC, all with AUC values of 1.00 (**Figure 3.3B**). Similarly, the lncRNAs metastasis associated lung adenocarcinoma transcript 1 (MALAT1) and LOC100190986, and the snoRNA SNORA11B (AUCs: 1.00) also presented a high accuracy for the identification of CCA *vs*. PSC (**Figure 3.3B**). Of note, several mRNA and non-coding RNAs provided excellent diagnostic values (AUC values up to 0.931 and 0.902, respectively) for the diagnosis of PSC, when compared with patients with UC and healthy individuals (**Figure 3.4**). Finally, general CCA transcript biomarkers (compared to PSC, UC and healthy individuals combined as one unique control group) were also identified and RFFL, zinc finger protein 266 (ZNF266) and OR4F3 constituted the mRNA transcripts with the highest AUC values (1.00, 0.976, and 0.960, respectively) while miR-551B, PMS2L4, and LOC643955 were the ncRNAs presenting the highest diagnostic capacity, displaying AUC values of 0.909, 0.880, and 0.873, respectively (**Figure 3.3C**).

3.3 Selective mRNAs Present in Serum EVs from Patients with CCA Mirror Their Levels in Human Tumor Tissue, CCA Cells In Vitro and EVs-Derived from Tumor Cholangiocytes

After identifying 2807 transcripts significantly altered in serum EVs from patients with CCA compared to patients with PSC, UC, and healthy individuals, we evaluated if the expression of these transcripts were also significantly changed in human CCA tissue, compared to non-tumor surrounding tissue, in two independent international cohorts of patients (TCGA and the “Copenhagen” cohorts). Importantly, 901 out of the 2807 selective RNA transcripts were also altered in the TCGA cohort, presenting the same trend of expression when compared to serum EVs, with 765 transcripts being upregulated while 136 transcripts were downregulated in comparison to non-tumor tissue (**Figure 3.5A, left**). These 765 transcripts were then cross-validated in the Copenhagen cohort, in which we were able to identify 479 shared transcripts with the same expression tendency, with 391 being upregulated and 88 reduced when compared with surrounding liver tissue (**Figure 3.5A, right**). After selecting the common mRNAs that share the same trend of expression in serum EVs and tumor tissue of patients with CCA compared to controls, we next evaluated their expression levels in two human CCA cell lines (EGI1 and TFK1) compared to NHCs in vitro, obtaining 156 commonly altered transcripts (**Figure 3.5B**). Finally, we isolated EVs from these two CCA cell lines and from NHCs and, after their characterization (**Figure 3.6**),[18] we evaluated their transcriptomic content

and cross-validated the previous candidate transcripts, resulting in 105 mRNAs with shared altered levels in serum EVs, tumor tissue, CCA cell lines, and in CCA-derived EVs (**Figure 3.5C**).

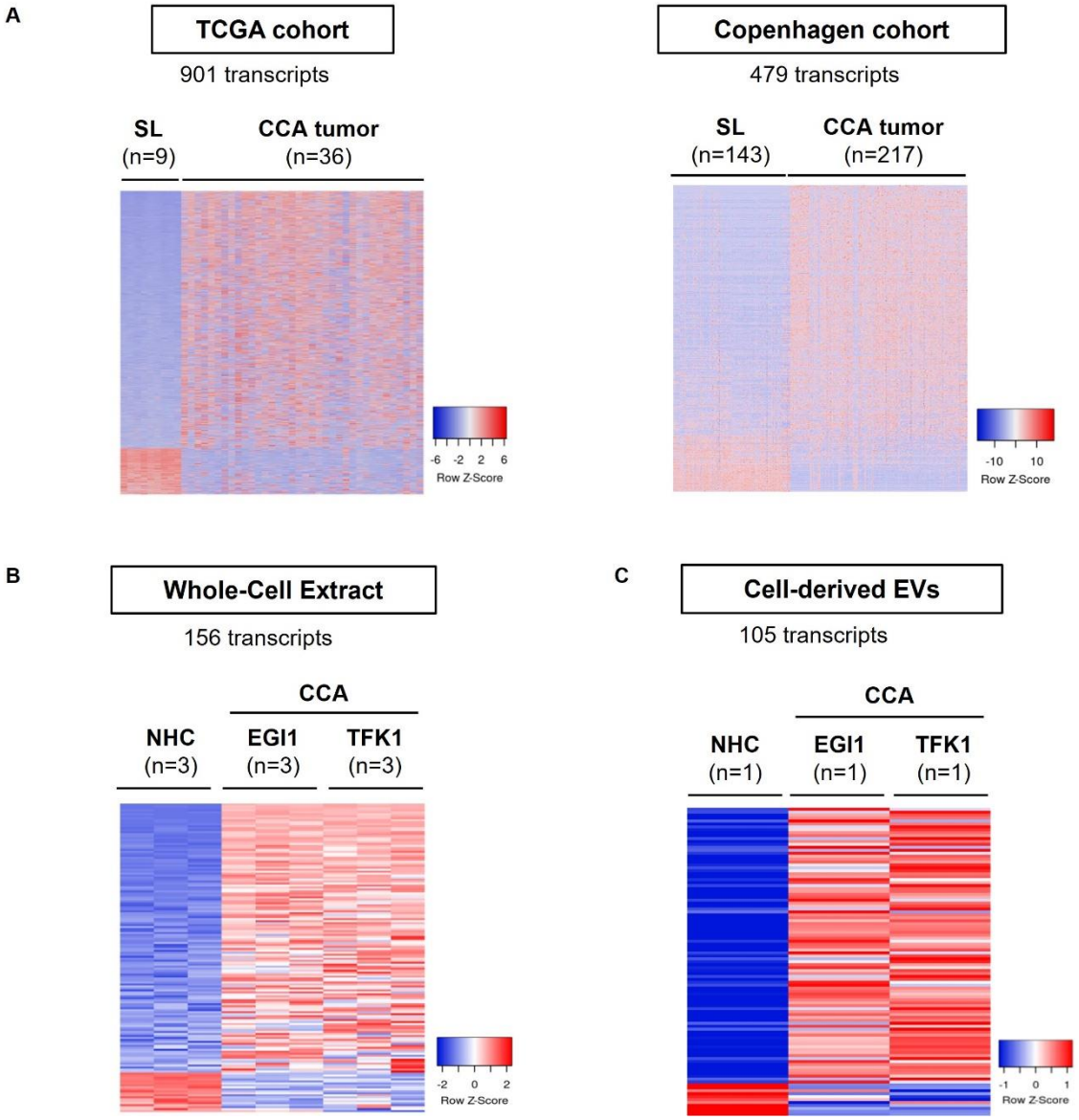


Figure 3.5. mRNAs commonly deregulated between serum EVs, CCA tumors from two independent cohorts of patients, tumor cells in vitro and in CCA-derived EVs. mRNAs differentially abundant in serum EVs from patients with CCA *vs*. (PSC + UC + Healthy individuals) were compared with the transcriptome of: (i) patients with CCA from TCGA (n = 36) and "Copenhagen" (n = 217) cohorts, (ii) CCA cells (EGI1 and TFK1) and cell-derived EVs compared to their respective control groups, further selecting the ones that are commonly expressed. Heatmap of the differentially expressed transcripts in (**A**) TCGA (left) and "Copenhagen" cohorts (right); (**B**) Whole-cell extracts from CCA cells and NHCs; (**C**) Cell-derived EVs; (**D**) Gene ontology (GO: FunRich database)[27] analysis of the 105 transcripts commonly altered in serum EVs, CCA human tumors, CCA cells and in cell-derived EVs, highlighting the biological processes and pathways in which the identified transcripts are involved, as well as their biological function. Abbreviations: CCA, cholangiocarcinoma, EVs, extracellular vesicles; NHCs, normal human cholangiocytes; SL, surrounding liver; TCGA, The cancer genome atlas.

D

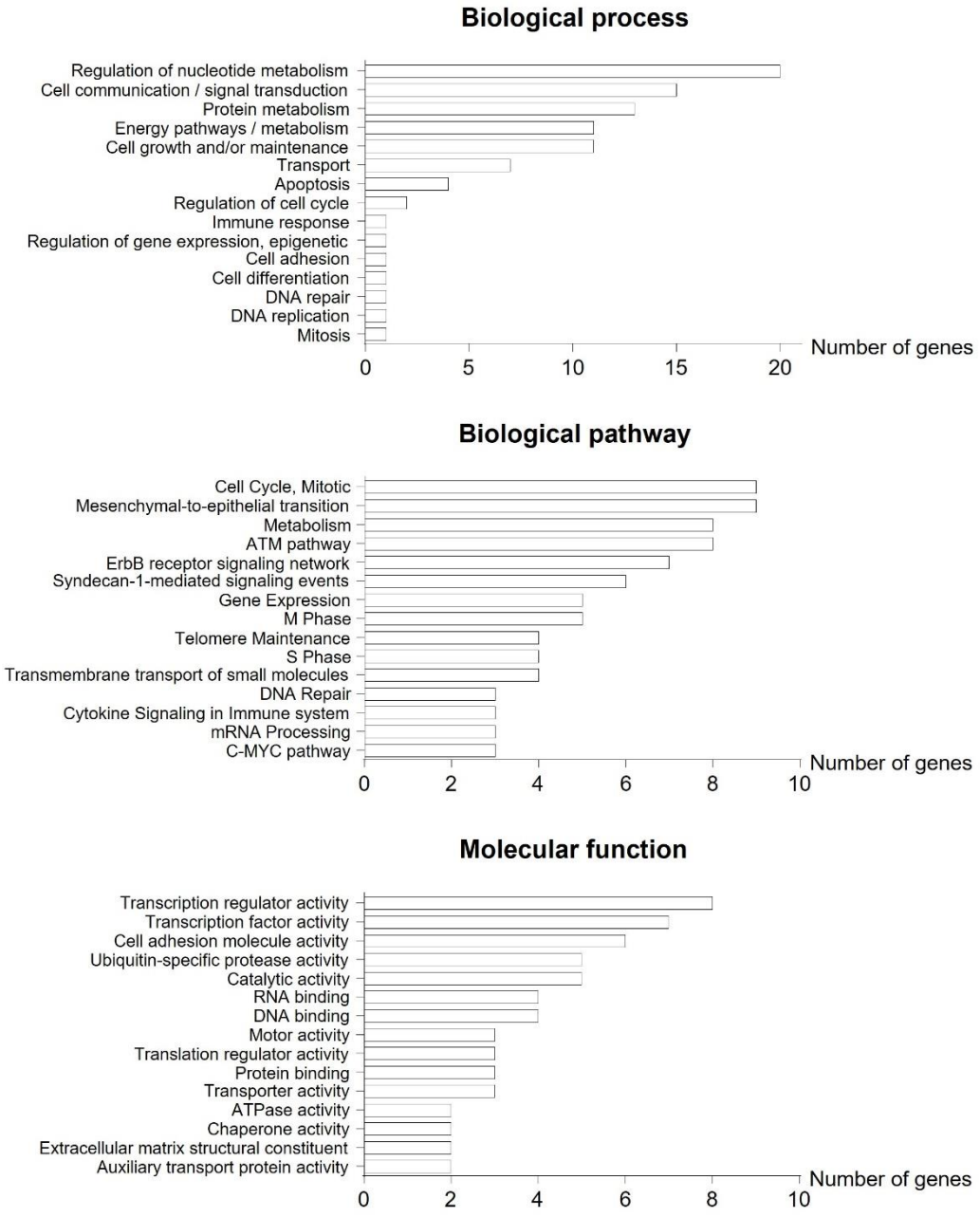
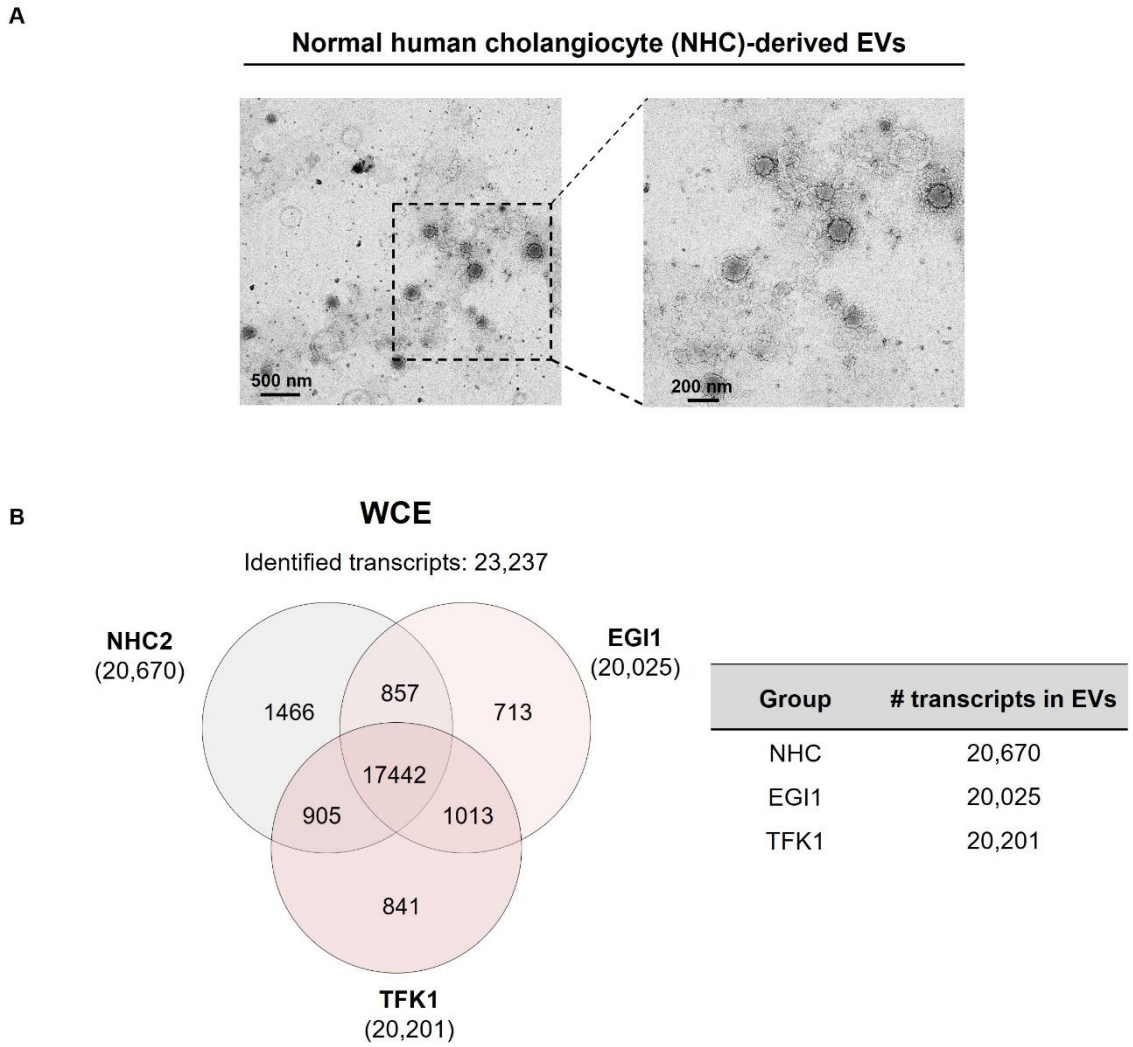


Figure 3.5. (Continued)



Group	# transcripts in EVs
NHC	20,670
EGI1	20,025
TFK1	20,201

Figure 3.6. Characterization of cholangiocyte-derived EVs transcriptome. In order to validate the protocol for EVs isolation from cholangiocytes, we used normal human cholangiocytes (NHCs). (**A**) TEM images of NHC-derived EVs showcasing the typical round shape (~150 nm) and morphology. (**B**) Venn diagrams showing the number of transcripts altered for each comparison and the total transcripts identified in each group.

Gene ontology (GO) analysis revealed that the 105 mRNAs previously identified code for proteins mainly involved in pivotal processes during carcinogenesis. Specifically, metabolic pathways (nucleic acid and protein metabolism), cell communication, signal transduction, energy, and cell growth/maintenance pathways constituted the most represented biological processes in which the identified mRNA transcripts are involved, therefore impacting in important cancer-promoting pathways, such as cell cycle regulation, mitosis, epithelial-mesenchymal transition (EMT), DNA repair, and telomere maintenance (**Figure 4D**). Of note, the expression of some of these transcripts positively correlated with worse disease severity (*i.e.*, advanced tumor stage and tumor dedifferentiation) in the TCGA and Copenhagen cohorts (**Figure 3.7**).

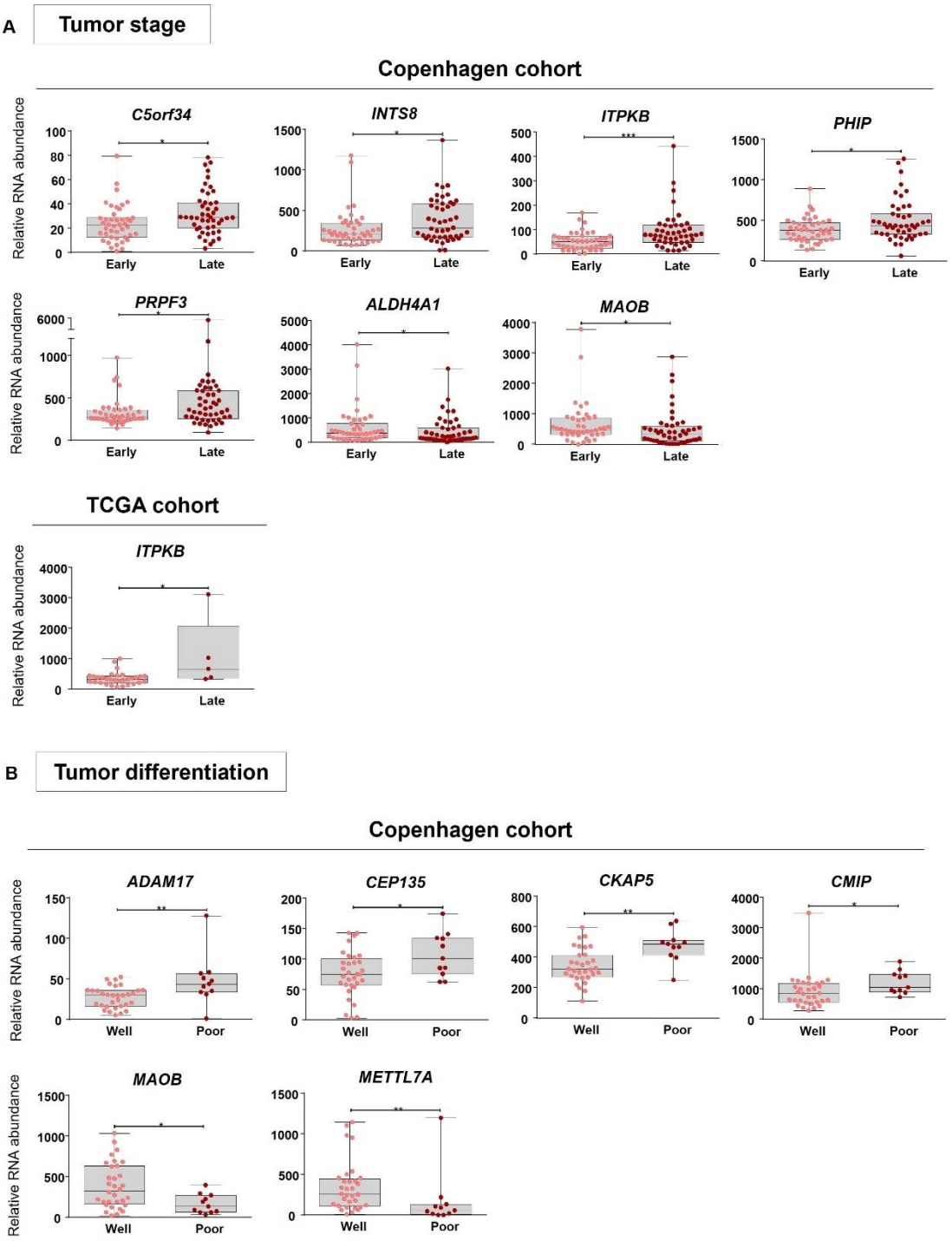


Figure 3.7. Comparative analysis of the levels of candidate serum liquid biopsy transcripts with CCA disease severity in the Copenhagen and TCGA cohorts. (**A**) Tumour stage. Abbreviations: Early, T1/2 tumor stage; Late, T3/4 tumor stage. (**B**) Tumour differentiation. Abbreviations: Well, well + well to moderate + moderate tumor differentiation; Poor, moderate to poor + poor differentiation.

Among all the 105 transcripts identified, c-Maf inducing protein (CMIP), glutamate decarboxylase 1 (GAD1), nucleoside diphosphate kinase 1 (NME1), CDP-diacylglycerol synthase 1 (CDS1), and cyclin-dependent kinases regulatory subunit 1 (CKS1B) constituted the best liquid biopsy candidate biomarkers, as their levels were increased in serum EVs and presented AUC values of 0.957, 0.928, 0.899, 0.893, and 0.891, respectively, for the diagnosis of CCA in comparison with patients with PSC, UC, and healthy individuals grouped together (**Figure 3.8A–E**). Importantly, a panel comprised of CMIP, NME1 and CKS1B provided the maximum diagnostic capacity (AUC: 1.000) for CCA in comparison with the group containing healthy individuals and patients with PSC and UC (**Figure 3.8F**). The mRNA levels of the aforementioned transcripts were also markedly increased in serum EVs from patients with CCA compared with the ones isolated from PSC patients, therefore potentially constituting novel biomarkers for the differential diagnosis of CCA and PSC. The mRNA expression levels of these genes were also increased in CCA tumor samples from the TCGA and Copenhagen cohorts, compared with either normal surrounding liver tissue or normal intrahepatic bile ducts, being also found upregulated in CCA tumor cells, compared to NHCs, and also in CCA-derived EVs (**Figure 3.8A–E**). Noteworthy, besides providing the best diagnostic value among all the liquid biopsy candidates, CMIP mRNA levels were found particularly increased in poorly-differentiated tumors in the Copenhagen cohort (**Figure 3.7B**).

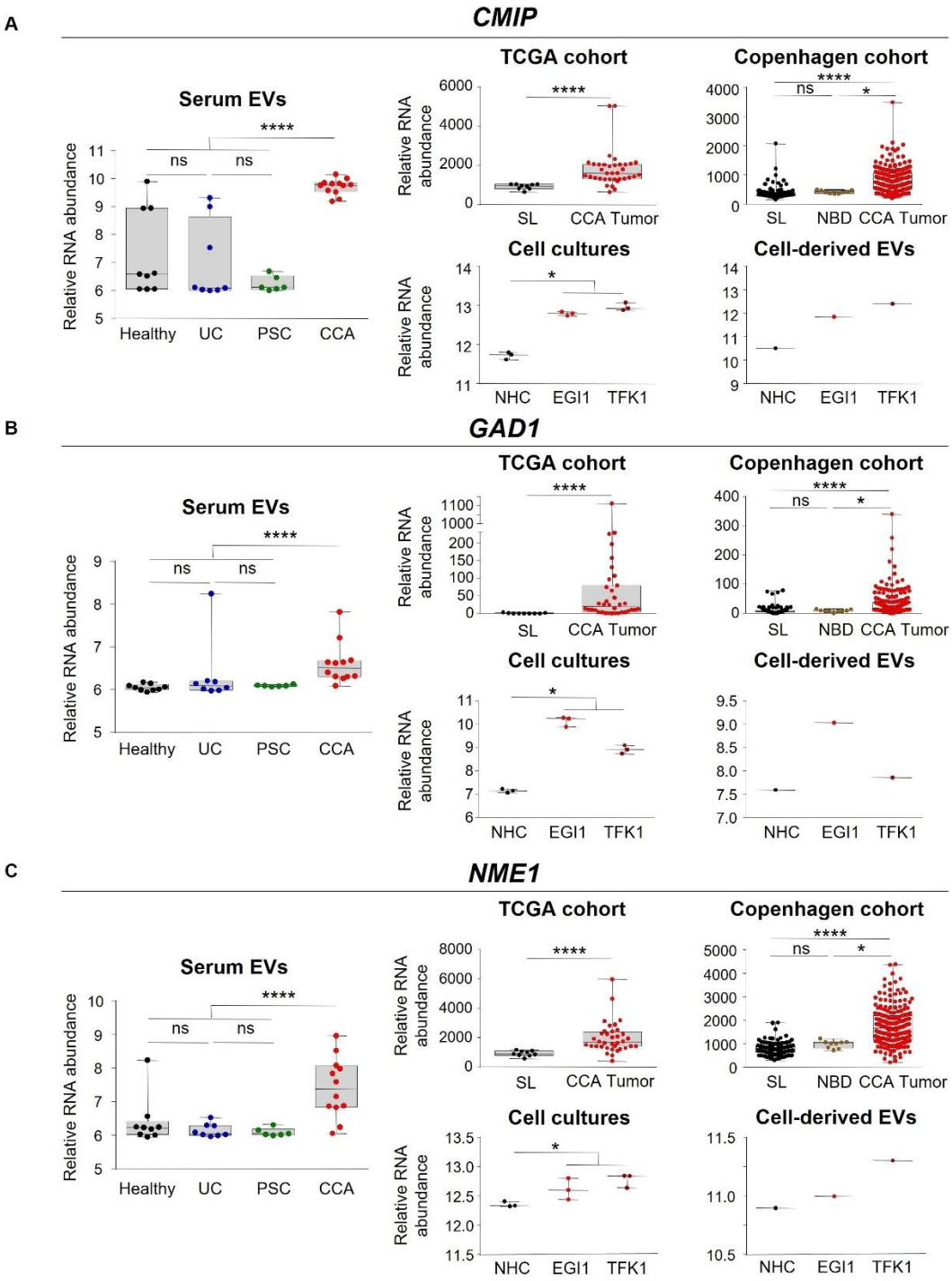
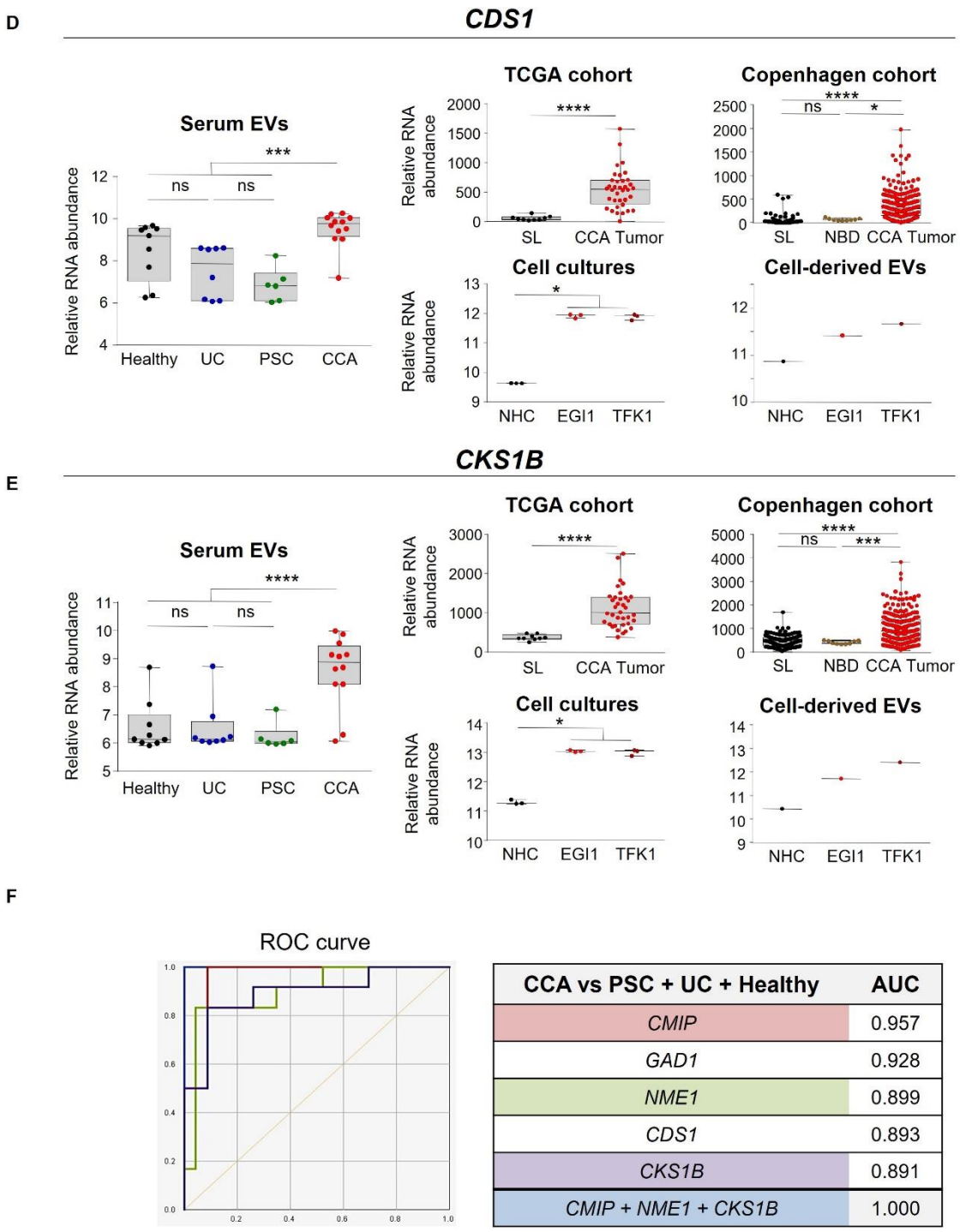


Figure 3.8. Selected liquid biopsy biomarkers for CCA. From the 105 mRNAs commonly altered in serum EVs, CCA human tumors, CCA cells and in cell-derived EVs compared to their corresponding controls, 5 biomarkers were selected based on their diagnostic capacity. Box plot diagrams with the mRNAs abundance in serum EVs (left), CCA tumors from the TCGA, and "Copenhagen" cohorts, CCA cells and cell-derived EVs (right), compared to their respective controls, for (**A**) c-Maf inducing protein (CMIP), (**B**) glutamate decarboxylase 1 (GAD1), and (**C**) NME/NM23 nucleoside diphosphate kinase 1 (NME1); (**D**) CDP-diacylglycerol synthase 1 (CDS1), and (**E**) CDC28 protein kinase regulatory subunit 1B (CKS1B). (**F**) Diagnostic prediction (ROC curves and AUC values) of the selected serum liquid biopsy biomarkers and for the combination of CMIP, NME1 and CKS1B for the diagnosis of CCA in comparison with (PSC + UC + Healthy individuals). Abbreviations: AUC, area under the receiver operating characteristic (ROC) curve; NHC, normal human cholangiocyte; NBD, normal bile ducts; SL, surrounding liver; TCGA, The cancer genome atlas.



CCA vs PSC + UC + Healthy	AUC
CMIP	0.957
GAD1	0.928
NME1	0.899
CDS1	0.893
CKS1B	0.891
CMIP + NME1 + CKS1B	1.000

Figure 3.8. (Continued)

3.4 Differential Transcriptomic Profiles of Urine EVs from Patients with CCA, PSC, UC, and Healthy Individuals

In urine EVs, a total of 11,323, 23,920, 22,159, and 26,066 RNAs were identified in healthy individuals and patients with UC, PSC, or CCA, respectively, accounting for a total of 27,319 transcripts detected (**Figure 3.9A**). Regarding the type of transcripts found, similar distribution was observed when compared with the profile obtained in serum EVs, abundantly detecting mRNAs, followed by ncRNAs such as pseudogenes and lncRNAs, miRNAs, and snoRNAs (**Figure 3.9B**). The analysis of urine EVs revealed differential transcriptomic profiles between patients with CCA, PSC, or UC, and healthy individuals.

In particular, 2386 transcripts were differentially identified in urine EVs from patients with CCA *vs*. healthy individuals, 1999 in CCA *vs*. PSC, and 1329 in CCA *vs*. other diseases (PSC and UC) and healthy individuals combined in a single group. In this regard, INO80 complex subunit D (INO80D), MAP6 domain containing 1 (MAP6D1) and Ras-related GTP binding D (RRAGD) constituted the most promising mRNA biomarkers in urine EVs isolated from patients with CCA, in comparison with healthy individuals, displaying AUC values of 1.000. Of note, the lncRNA HLA complex group 4 (HCG4), miR200c, and the lncRNA LOC100134868 also presented high accuracies for the diagnosis of CCA, with AUC values of 0.930, 0.904, and 0.896, respectively (**Figure 3.10A**). Regarding the differential identification of CCA and PSC, the mRNA transcripts CAP-Gly domain containing linker protein 3 (CLIP3), Vascular cell adhesion molecule 1 (VCAM1) and Tripartite motif containing 33 (TRIM33) displayed excellent AUC values (0.965), in parallel with the pseudogene ATP synthase F1 subunit epsilon pseudogene 2 (ATP5EP2), the lncRNA LOC100134713, and the Small nucleolar RNA, H/ACA box 8 (SNORA8) (AUC: 0.939, 0.930, and 0.922, respectively), thus allowing to distinguish patients with CCA and PSC with high sensitivity and specificity values (**Figure 3.10B**). When comparing urine EVs from patients with PSC *vs*. UC and healthy individuals considered as a unique group, several mRNA and non-coding RNAs arose as potential biomarkers for the diagnosis of PSC, with AUC values up to 1.000 and 0.965, respectively (**Figure 3.11**). Finally, by comparing CCA to the group containing patients with other diseases (PSC and UC) and healthy individuals, the mRNA metallothionein 1F (MT1F), glutathione peroxidase 3 (GPX3), and lactate dehydrogenase A (LDHA) stood out as the ones with the best AUC values (0.915, 0.897, and 0.894, respectively), along with the ncRNAs U11 small nuclear (RNU11), LOC257358, and vault RNA 1-1 (VTRNA1-1) (0.830, 0.812, and 0.777, respectively) (**Figure 3.10C**), therefore constituting promising urine EV biomarkers for the accurate diagnosis of CCA.

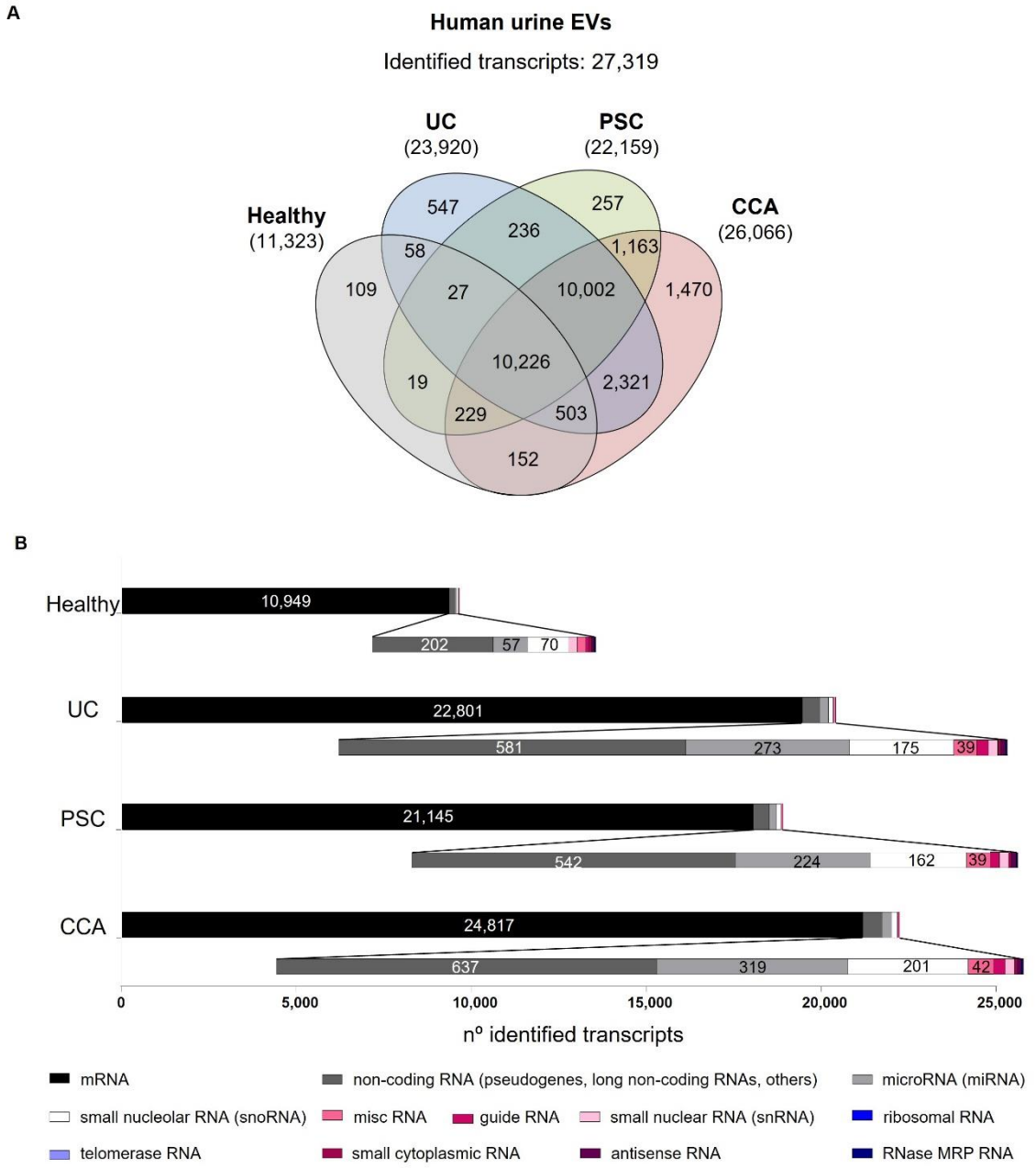
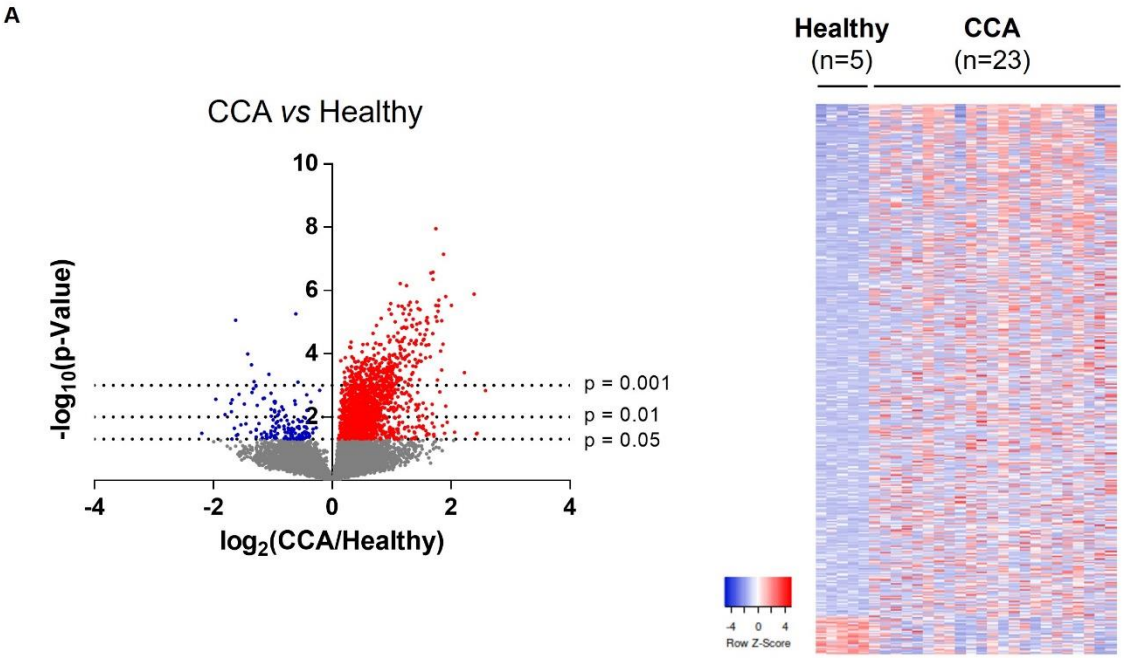


Figure 3.9. Comparative transcriptomic analysis of urine EVs from patients with CCA, PSC, or UC, and healthy individuals. (**A**) Venn diagrams showing the number of transcripts identified per group. (**B**) Number of transcripts identified within each study group, subclassified according to their type [messenger RNA (mRNA), non-coding RNA (including mostly pseudogenes and long non-coding RNAs, among others), miscelaneous RNA (miscRNA), guide RNA, microRNA (miRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), ribosomal RNA, telomerase RNA, small cytoplasmic RNA, antisense RNA and RNase MRP RNA]. In all groups, mRNAs constitute the most abundantly identified RNAs, followed by non-coding RNAs (pseudogenes, lncRNAs, and others), miRNAs, and snoRNAs.

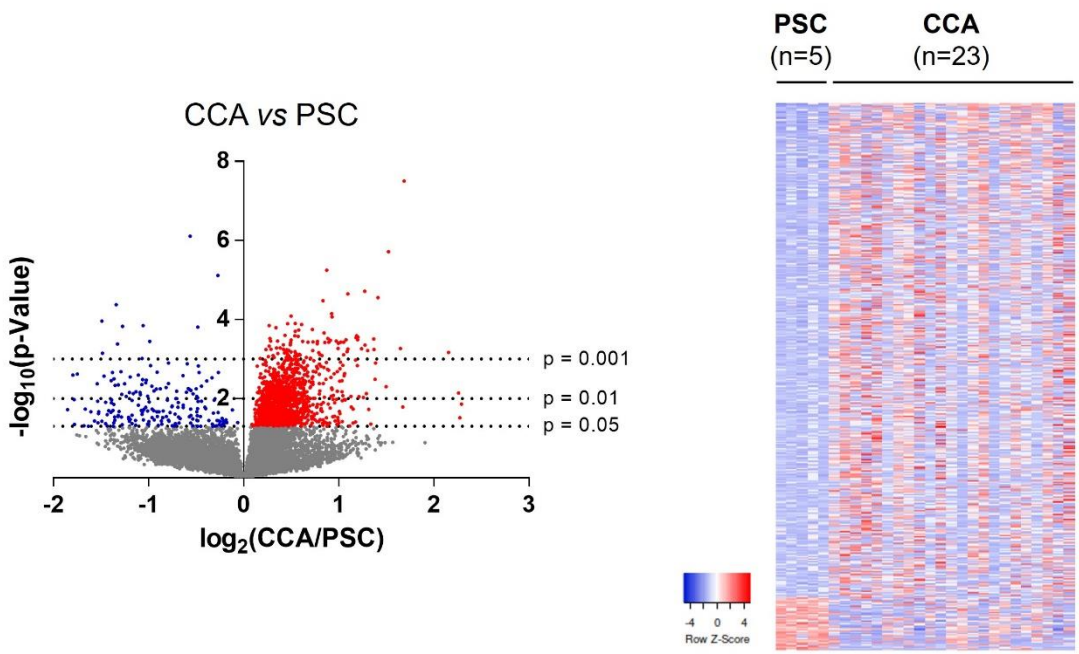


Transcript		CCA (n=23) *vs* Healthy (n=5)								
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
INO80D	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
MAP6D1	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
RRAGD	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
PLCE1	mRNA	**0.991** (0.964, 1.000)	<0.001	95.7	100.0	100.0	83.3	>100	0.04	96.4
ERRFI1	mRNA	**0.983** (0.940, 1.000)	<0.001	95.7	100.0	100.0	83.3	>100	0.04	96.4
BCAP31	mRNA	**0.983** (0.940, 1.000)	<0.001	95.7	100.0	100.0	83.3	>100	0.04	96.4
PPIG	mRNA	**0.983** (0.940, 1.000)	<0.001	95.7	100.0	100.0	83.3	>100	0.04	96.4
METTL2B	mRNA	**0.983** (0.939, 1.000)	<0.001	91.3	100.0	100.0	71.4	>100	0.09	92.9
TOM1L2	mRNA	**0.965** (0.901, 1.000)	<0.01	91.3	100.0	100.0	71.4	>100	0.09	92.9
APITD1	mRNA	**0.965** (0.900, 1.000)	<0.01	87.0	100.0	100.0	62.5	>100	0.13	89.3

Transcript		CCA (n=23) *vs* Healthy (n=5)								
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
HCG4	lncRNA	**0.930** (0.835, 1.000)	<0.01	87.0	100.0	100.0	62.5	>100	0.13	89.3
MIR200C	miRNA	**0.904** (0.743, 1.000)	<0.01	95.7	80.0	95.7	80.0	4.78	0.05	92.9
LOC100134868	lncRNA	**0.896** (0.775, 1.000)	<0.001	87.0	100.0	100.0	62.5	>100	0.13	89.3
TTC3L	pseudogene	**0.896** (0.771, 1.000)	<0.01	73.9	100.0	100.0	45.5	>100	0.26	78.6
MIR660	miRNA	**0.887** (0.763, 1.000)	<0.01	78.3	100.0	100.0	50.0	>100	0.22	82.1

Figure 3.10. Differential transcriptomic profile of urine EVs and diagnostic capacity. Volcano plot (−log10(p-value) and log2(fold-change); up left), heatmap of the differentially expressed transcripts (up right) and diagrams with the diagnostic capacity with the highest AUC values of the 10 selected mRNAs and 5 selected non-coding RNAs in serum EVs from (**A**) CCA *vs*. Healthy individuals; (**B**) CCA *vs*. PSC; (**C**) CCA *vs*. (PSC + UC + Healthy individuals). Abbreviations: AI, accuracy index; AUC, area under the receiver operating characteristic curve; CI, confidence interval; miRNA, microRNA; lncRNA, long non-coding RNA; miscRNA, miscellaneous RNA; NLR, negative likelihood ratio; NPV, negative predictive value; PLR, positive likelihood ratio; PPV, positive predictive value; SEN, sensitivity; snRNA, small nuclear RNA; snoRNA, small nucleolar RNA; SPE, specificity; vtRNA, vault RNA.

B

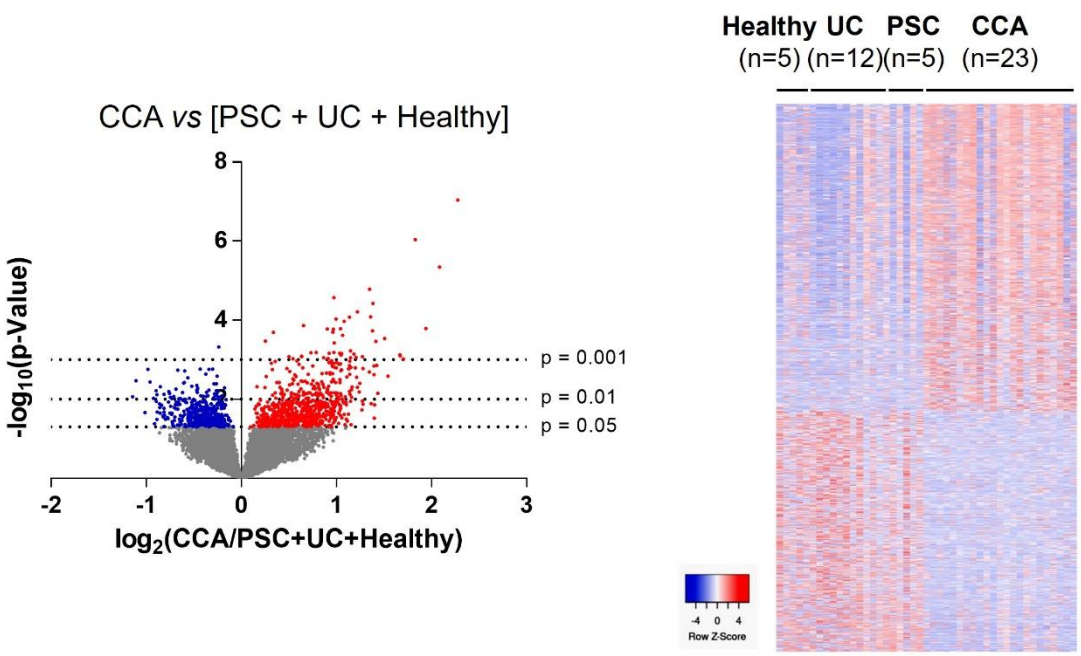


Transcript		CCA (n=23) *vs* PSC (n=5)							
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
CLIP3	mRNA	**0.965** (0.900, 1.000)	<0.001	87.0	100.0	100.0	62.5	>100	0.13	89.3
VCAM1	mRNA	**0.965** (0.900, 1.000)	<0.001	87.0	100.0	100.0	62.5	>100	0.13	89.3
TRIM33	mRNA	**0.965** (0.900, 1.000)	<0.001	87.0	100.0	100.0	62.5	>100	0.13	89.3
POLR3A	mRNA	**0.965** (0.895, 1.000)	<0.001	95.7	100.0	100.0	83.3	>100	0.04	96.4
ADAMTS6	mRNA	**0.965** (0.891, 1.000)	<0.001	82.6	100.0	100.0	55.6	>100	0.17	85.7
C17orf56	mRNA	**0.965** (0.891, 1.000)	<0.001	82.6	100.0	100.0	55.6	>100	0.17	85.7
INSM1	mRNA	**0.957** (0.883, 1.000)	<0.01	87.0	100.0	100.0	62.5	>100	0.13	89.3
MTHFR	mRNA	**0.948** (0.868, 1.000)	<0.01	87.0	100.0	100.0	62.5	>100	0.13	89.3
HSPE1	mRNA	**0.948** (0.864, 1.000)	<0.01	82.6	100.0	100.0	55.6	>100	0.17	85.7
AOF2	mRNA	**0.948** (0.861, 1.000)	<0.01	91.3	100.0	100.0	71.4	>100	0.09	92.9

Transcript		CCA (n=23) *vs* PSC (n=5)								
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
ATP5EP2	pseudogene	**0.939** (0.851, 1.000)	<0.01	87.0	100.0	100.0	62.5	>100	0.13	89.3
LOC100134713	lncRNA	**0.930** (0.828, 1.000)	<0.01	82.6	100.0	100.0	55.6	>100	0.17	85.7
SNORA8	snoRNA	**0.922** (0.820, 1.000)	<0.01	82.6	100.0	100.0	55.6	>100	0.17	85.7
MIR183	miRNA	**0.896** (0.765, 1.000)	<0.01	91.3	80.0	95.5	66.7	4.57	0.11	89.3
SNORA2B	snoRNA	**0.887** (0.760, 1.000)	<0.01	73.9	100.0	100.0	45.5	>100	0.26	78.6

Figure 3.10. (Continued)

C

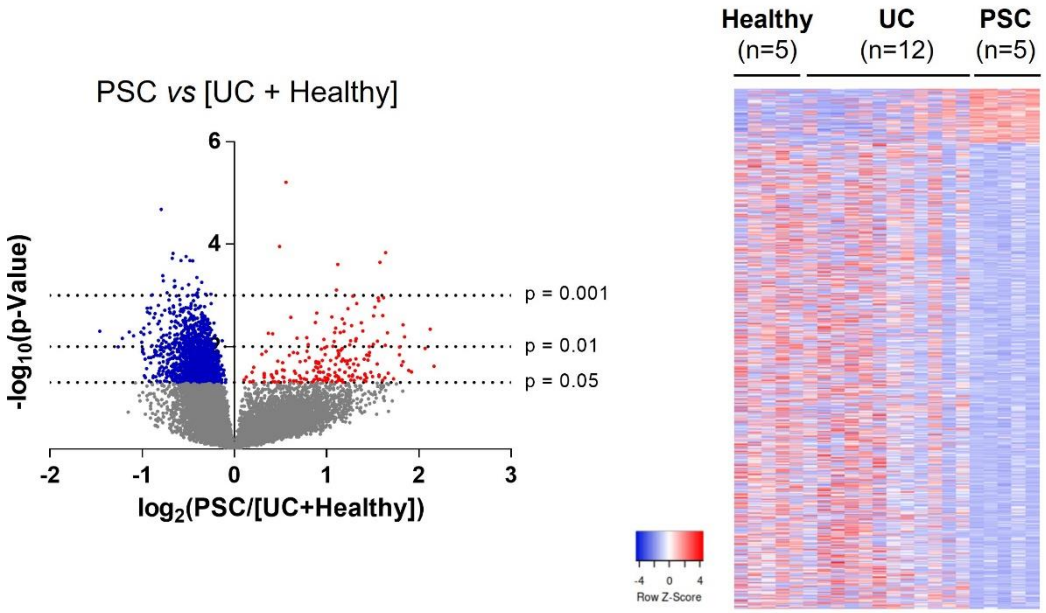


Transcript		CCA (n=23) *vs* [PSC + UC + Healthy] (n=22)								
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
MT1F	mRNA	**0.915** (0.826, 1.000)	<0.001	73.9	100.0	100.0	78.6	>100	0.26	86.7
GPX3	mRNA	**0.897** (0.789, 1.000)	<0.001	82.6	95.5	95.0	84.0	18.17	0.18	88.9
LDHA	mRNA	**0.894** (0.797, 0.992)	<0.001	87.0	86.4	87.0	86.4	6.38	0.15	86.67
MT1H	mRNA	**0.874** (0.752, 0.995)	<0.001	87.0	90.9	90.9	87.0	9.57	0.14	88.9
MT2A	mRNA	**0.864** (0.752, 0.975)	<0.001	78.3	90.9	90.0	80.0	8.61	0.24	84.4
SUCLG1	mRNA	**0.864** (0.749, 0.978)	<0.001	69.6	95.5	94.1	75.0	15.3	0.32	82.2
MT1G	mRNA	**0.856** (0.736, 0.975)	<0.001	69.6	100.0	100.0	75.9	>100	0.30	84.4
TPCN1	mRNA	**0.854** (0.739, 0.969)	<0.001	73.9	86.4	85.0	76.0	5.42	0.30	80.0
ALDOB	mRNA	**0.849** (0.731, 0.967)	<0.001	78.3	86.4	85.7	79.2	5.74	0.25	88.2
ODC1	mRNA	**0.840** (0.716, 0.964)	<0.001	82.6	81.8	82.6	81.8	4.54	0.21	88.2

Transcript		CCA (n=23) *vs* Others (n=22)								
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
RNU11	snRNA	**0.830** (0.711, 0.949)	<0.001	56.5	100.0	100.0	68.8	>100	0.43	77.8
LOC257358	miscRNA	**0.812** (0.690, 0.934)	<0.001	52.2	95.5	92.3	65.6	11.48	0.50	73.3
VTRNA1-1	vtRNA	**0.777** (0.641, 0.912)	<0.01	73.9	72.7	73.9	72.7	2.71	0.36	73.3
AURKAPS1	pseudogene	**0.771** (0.630, 0.911)	<0.01	87.0	68.2	74.1	83.3	2.73	0.19	77.8
MIR483	miRNA	**0.763** (0.625, 0.900)	<0.01	87.0	54.5	66.7	80.0	1.91	0.24	71.1

Figure 3.10. (Continued)

A



Transcript		PSC (n=5) *vs* [UC + Healthy] (n=17)								
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
ZSCAN2	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
LRRC8D	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
DDX54	mRNA	**1.000** (1.000, 1.000)	<0.001	100.0	100.0	100.0	100.0	>100	0.00	100.0
DLGAP3	mRNA	**0.988** (0.952, 1.000)	<0.001	100.0	94.1	83.3	100.0	17.00	0.00	95.5
ELOVL6	mRNA	**0.988** (0.952, 1.000)	<0.001	100.0	94.1	83.3	100.0	17.00	0.00	95.5
KIAA0195	mRNA	**0.988** (0.952, 1.000)	<0.001	100.0	94.1	83.3	100.0	17.00	0.00	95.5
SEMA3F	mRNA	**0.976** (0.920, 1.000)	<0.001	100.0	94.1	83.3	100.0	17.00	0.00	95.5
BCL2L1	mRNA	**0.976** (0.918, 1.000)	<0.001	100.0	88.2	71.4	100.0	8.50	0.00	90.9
PDZRN4	mRNA	**0.965** (0.889, 1.000)	<0.01	100.0	94.1	83.3	100.0	17.00	0.00	95.5
FOXP1	mRNA	**0.965** (0.889, 1.000)	<0.01	100.0	94.1	83.3	100.0	17.00	0.00	95.5

Transcript		PSC (n=5) *vs* [UC + Healthy] (n=17)								
Symbol	Transcript type	**AUC** (95% CI)	**p-Value**	**SEN** (%)	**SPE** (%)	**PPV** (%)	**NPV** (%)	**PLR**	**NLR**	**AI** (%)
SNORD85	snoRNA	**0.965** (0.886, 1.000)	<0.01	100.0	82.4	62.5	100.0	5.67	0.00	86.4
LOC286135	lincRNA	**0.941** (0.843, 1.000)	<0.01	100.0	82.4	62.5	100.0	5.67	0.00	86.4
MIR34B	miRNA	**0.906** (0.780, 1.000)	<0.01	100.0	82.4	62.5	100.0	5.67	0.00	86.4
LOH12CR2	lncRNA	**0.906** (0.779, 1.000)	<0.01	100.0	82.4	62.5	100.0	5.67	0.00	86.4
MIR183	miRNA	**0.906** (0.762, 1.000)	<0.01	80.0	94.1	80.0	94.1	13.60	0.21	90.9

Figure 3.11. Differential transcriptomic profile and diagnostic efficacy of urine EV transcripts. (**A**) Volcano plot [-log10(p-value) and log2(fold-change); up left], heatmap of the differentially expressed transcripts (up right) and diagrams with the diagnostic capacity with the highest AUC values of the 10 selected mRNAs and 5 selected non-coding RNAs in serum EVs from PSC vs [UC + Healthy individuals]. (**B**) Potential serum EV biomarkers for PSC. Abbreviations: AI, accuracy index; AUC, area under the receiver operating characteristic curve; CI, confidence interval; lncRNA, long non-coding RNA; miRNA, microRNA; NLR, negative likelihood ratio; NPV, negative predictive value; PLR, positive likelihood ratio; PPV, positive predictive value; SEN, sensitivity; snoRNA, small nucleolar RNA; SPE, specificity.

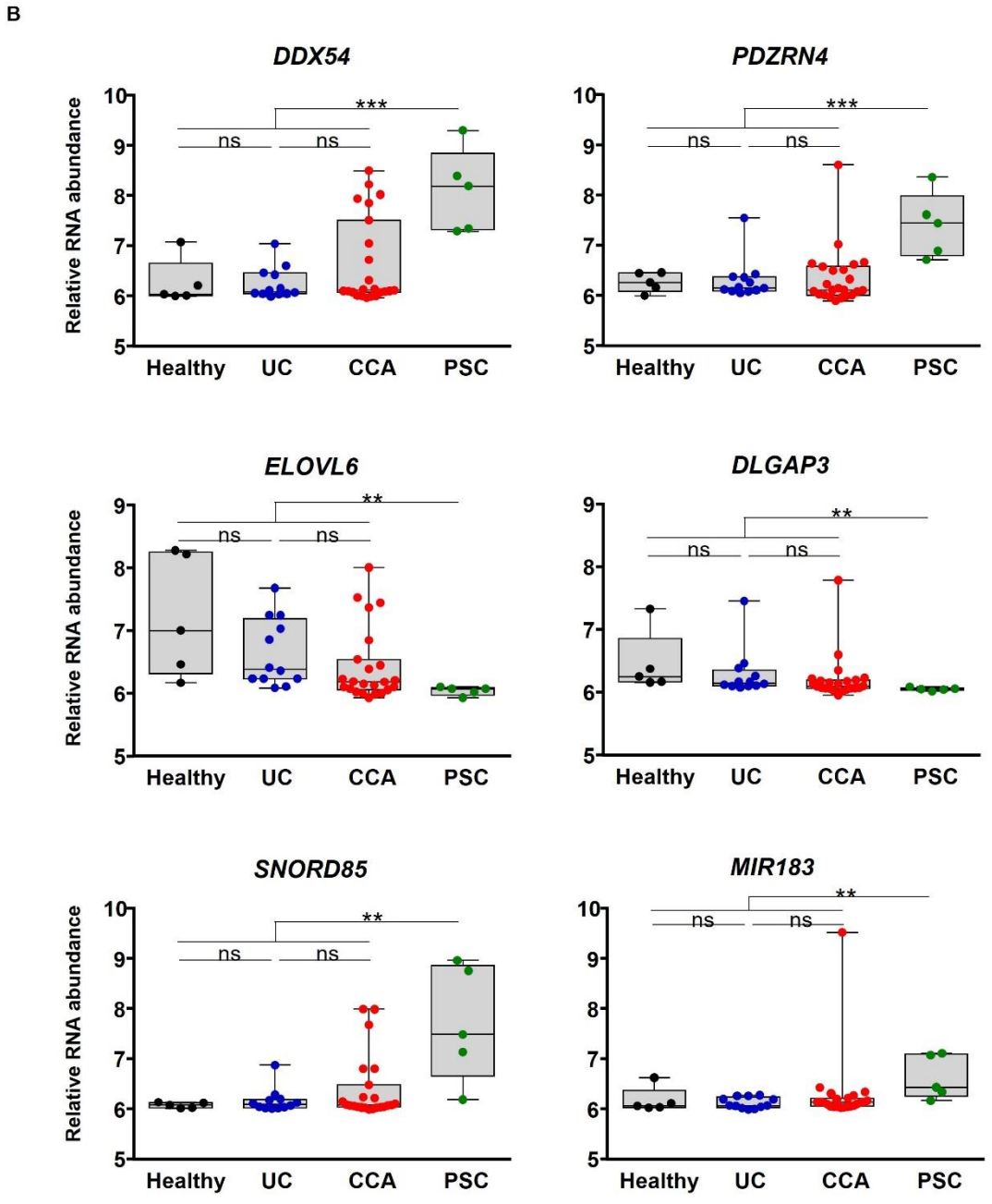


Figure 3.11. (Continued)

3.5 Selective mRNAs Present in Urine EVs from Patients with CCA Mimic Their Levels in Human Tumor Tissue, CCA Cells In Vitro, and EVs-Derived from Tumor Cholangiocytes

Similar to our previous analysis on the serum EV biomarkers identified in patients with CCA, compared to all the other study groups, we now selected the 1329 RNA transcripts that were significantly altered in urine EVs from patients with CCA and compared their expression levels with CCA and surrounding tumor tissues from both TCGA and Copenhagen cohorts. After performing a comprehensive analysis of these mRNAs in the TCGA cohort, we were able to identify 390 dysregulated transcripts (305 upregulated and 85 downregulated) that are commonly altered in urine EVs and in tumor samples from patients with CCA (**Figure 3.12A, left**). Additionally, compared to the changes observed in urine EVs, 259 transcripts also shared the same pattern of alteration in the Copenhagen cohort, with 206 transcripts presenting increased expression while 53 transcripts were reduced, when compared with surrounding liver (**Figure 3.12A, right**). The comparison of these 259 transcripts with the differential transcriptome of CCA cell lines compared to NHCs, revealed 84 shared mRNAs that were altered in CCA cells, with 69 being upregulated and 15 displaying decreased expression (**Figure 3.12B**). In EVs isolated from CCA and NHC cell cultures, we were able to identify 39 mRNAs (34 upregulated and 5 downregulated) commonly altered with urine EVs, tumor tissue, and CCA cells (**Figure 3.12C**).

In order to evaluate the role of these transcripts in carcinogenesis, we conducted a GO analysis and observed that, in resemblance with what we previously found in serum EVs, these mRNA transcripts code for proteins that are predominantly related with tumor development and progression, namely metabolic pathways (nucleic acids and protein metabolism), signal transduction, EMT and immune response. Although less represented, some transcripts were also linked to energy and cell growth/maintenance pathways (**Figure 3.12D**).

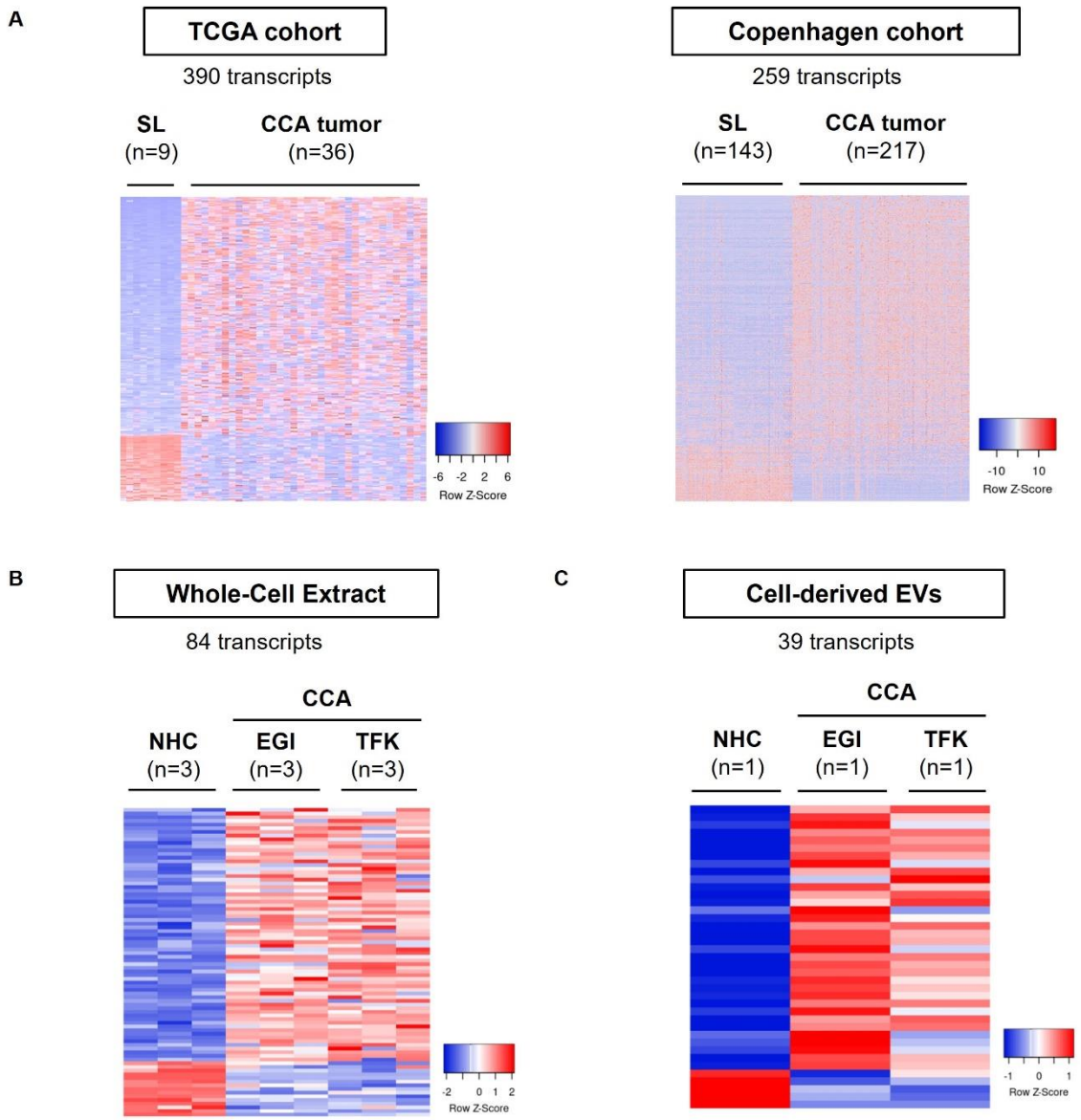


Figure 3.12. mRNAs commonly deregulated between urine EVs, CCA tumors from two independent cohorts of patients, tumor cells in vitro and in CCA-derived EVs. mRNAs differentially abundant in serum EVs from patients with CCA *vs*. (PSC + UC + Healthy individuals) were compared with the transcriptome of: (i) patients with CCA from The Cancer Genome Atlas TCGA (n = 36) and "Copenhagen" (n = 217) cohorts, (ii) CCA cells (EGI1 and TFK1) and cell-derived EVs compared to their respective control groups, further selecting the ones that are commonly expressed. Heatmap of the differentially expressed transcripts in (**A**) TCGA (left) and "Copenhagen" cohorts (right); (**B**) Whole-cell extracts from CCA cells and normal human cholangiocytes (NHCs); (**C**) Cell-derived EVs; (**D**) Gene ontology (GO: FunRich database)[27] analysis of the 105 transcripts commonly altered in serum EVs, CCA human tumors, CCA cells and in cell-derived EVs, highlighting the biological processes and pathways in which the identified transcripts are involved, as well as their biological function. Abbreviations: EVs, extracellular vesicles; NHC, normal human cholangiocyte; SL, surrounding liver; TCGA, The cancer genome atlas.

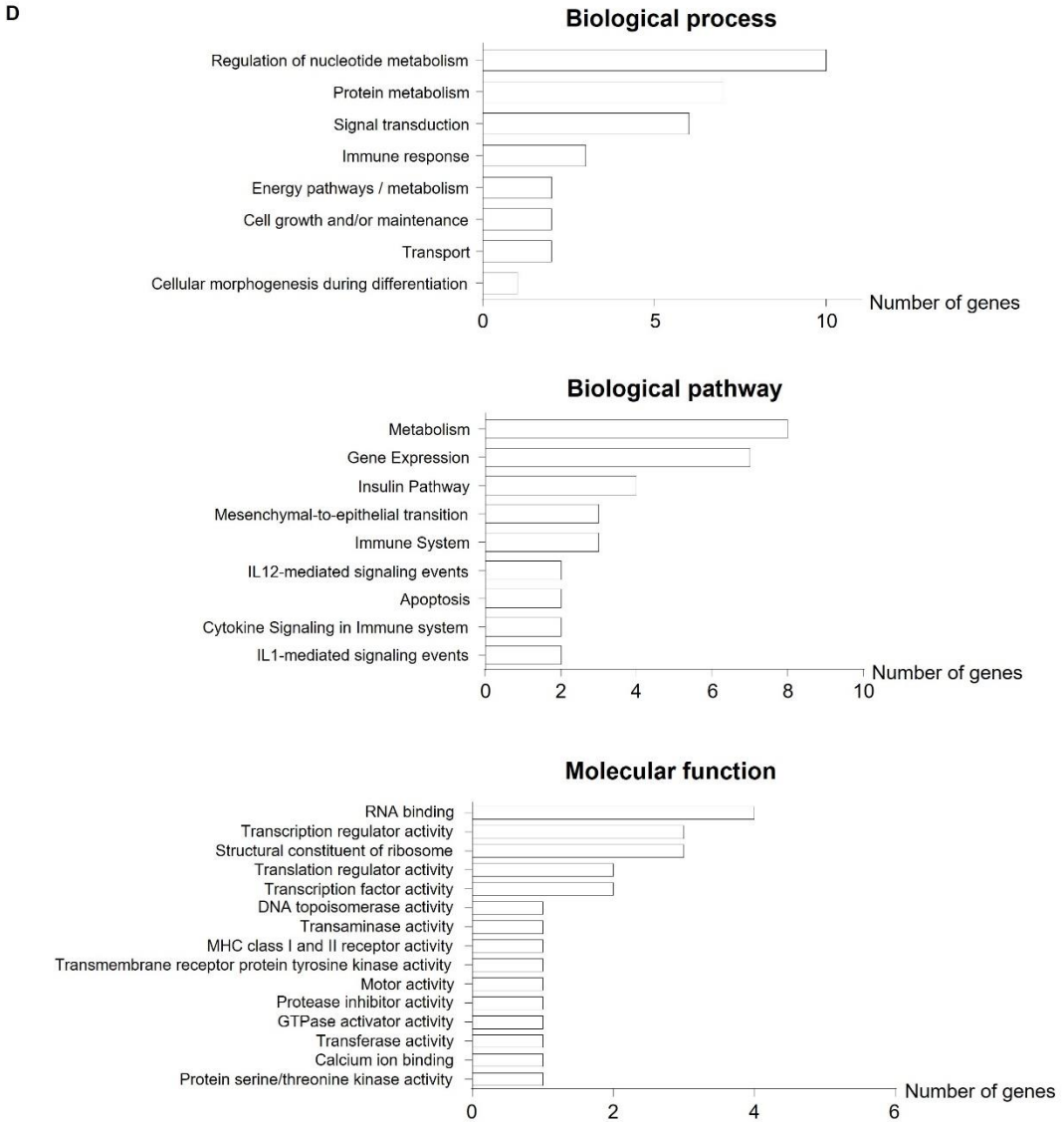


Figure 3.12. (Continued)

In this regard, the transcripts ubiquitin conjugating enzyme E2 C (UBE2C) and serine protease inhibitor B1 (SERPINB1) arose as potential liquid biopsy biomarkers, being increased in urine EVs isolated from patients with CCA in comparison to a group containing patients with PSC, UC and healthy controls (**Figure 3.13A,B**). Combining these urine biomarkers into one panel increased their diagnostic accuracy, providing an AUC value of 0.812 for the diagnosis of CCA (**Figure 3.13C**). Noteworthy, the expression levels of these transcripts were also markedly upregulated in CCA tumor samples from the TCGA and Copenhagen cohorts, when compared with both normal surrounding liver specimens and/or normal intrahepatic bile ducts, presenting also increased expression in CCA cells and in CCA-derived EVs, when compared with NHCs (**Figure 3.13A,B**). Importantly, SERPINB1 mRNA levels increased with disease severity in the Copenhagen cohort, being particularly overexpressed in advanced tumor stages compared with early stage CCAs (**Figure 3.14A**). Furthermore, although not being presented as one of the best liquid biopsy candidate, Tctex1 domain containing 2

(TCTEX1D2) levels were found upregulated in poorly-differentiated tumors compared with well-differentiated ones (**Figure 3.14B**).

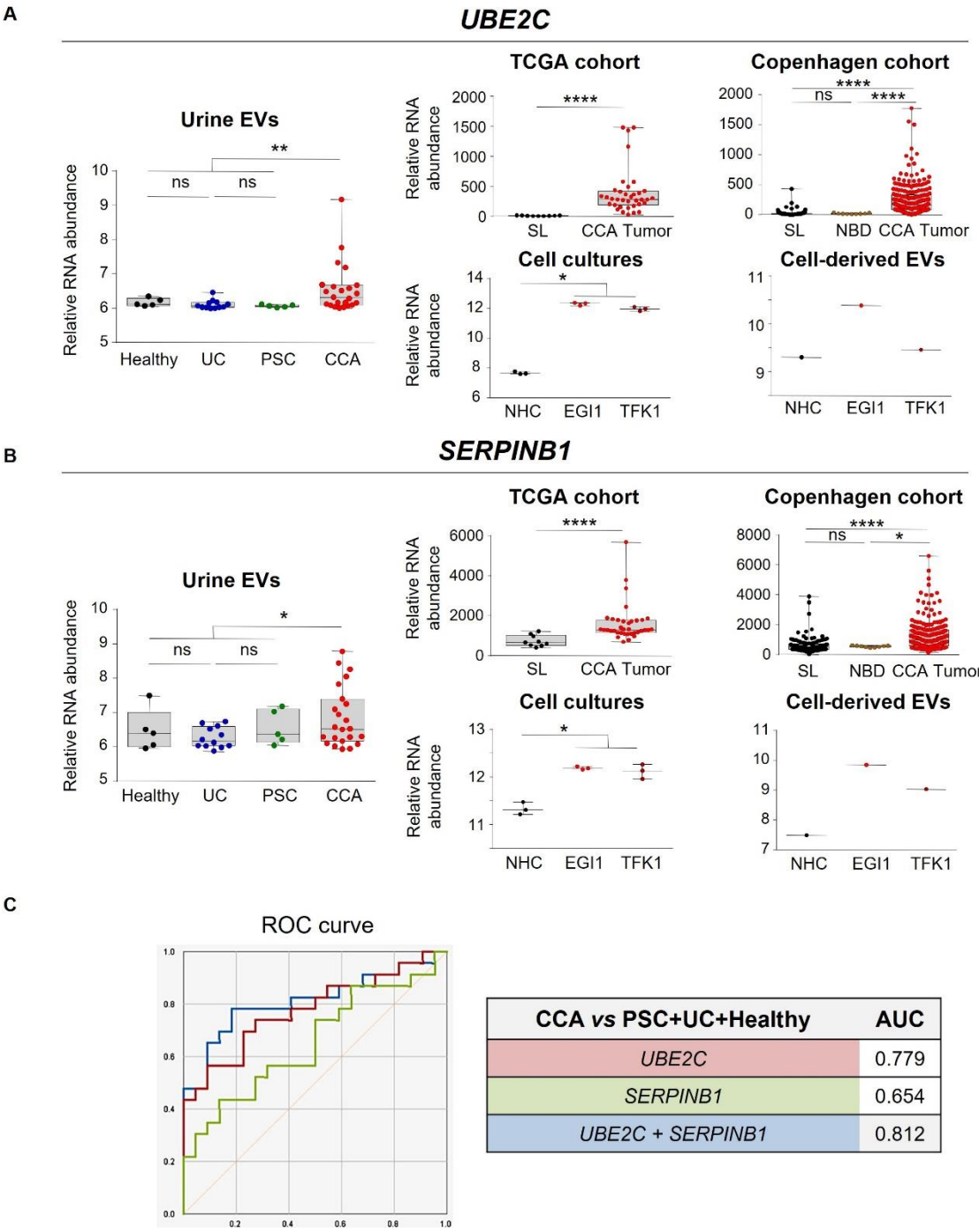


Figure 3.13. Potential urine liquid biopsy markers for CCA. From the 39 transcripts commonly found in serum EVs and differentially expressed in patient samples, CCA cells and in cell line-derived EVs, 2 potential urine liquid biopsy markers with the best diagnostic capacity were selected. Box plot diagrams with the mRNA transcript abundance in urine EVs (left) and the expression in the TCGA and "Copenhagen" cohorts, cholangiocyte cell lines and cell line-derived EVs (right) for (**A**) Ubiquitin conjugatin enzyme E2 C (UBE2C) and (**B**) Serine proteinase inhibitor B1 (SERPINB1). (**C**) Diagnostic prediction (ROC curves and AUC values) of the selected urine liquid biopsy markers and from the combination of UBE2C and SERPINB1 for the diagnosis of CCA in comparison with (PSC + UC + Healthy individuals). Abbreviations: AUC, area under the receiver operating characteristic curve; EVs, extracellular vesicles; NHC, normal human cholangiocyte; NBD, normal bile ducts; SL, surrounding liver; TCGA, The cancer genome atlas.

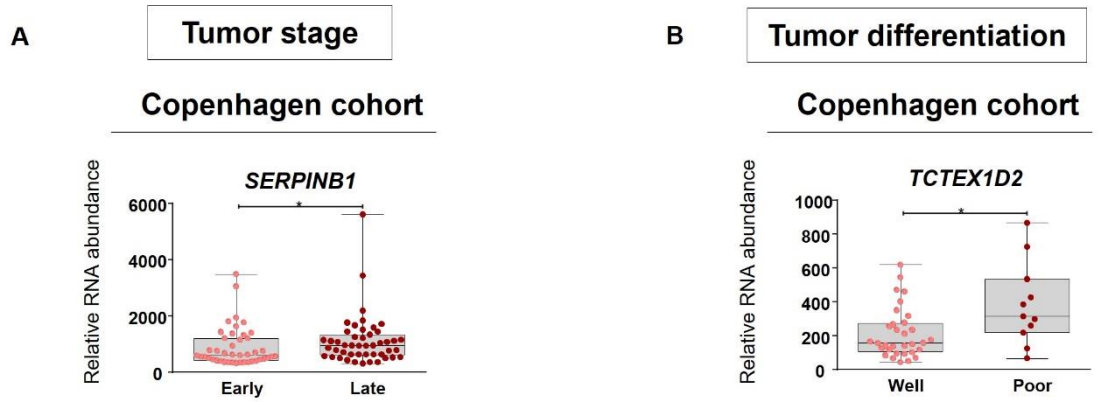


Figure 3.14. Comparative analysis of the levels of candidate urine liquid biopsy transcripts with CCA disease severity in the Copenhagen and TCGA cohorts. (**A**) Tumour stage; (**B**) Tumour differentiation. Abbreviations: Early, T1/2 tumor stage; Late, T3/4 tumor stage; Well, well + well to moderate + moderate tumor differentiation; Poor, moderate to poor + poor differentiation.

4. DISCUSSION

In the last decade, a considerable effort has been made to identify novel non-invasive biomarkers for the early and accurate diagnosis of CCA.[7] Here, we report for the first time the differential RNA profile of serum and urine EVs from patients with CCA, PSC, or UC, and healthy individuals, identifying new potential biomarkers with high diagnostic capacity. Noteworthy, some of the altered mRNAs were similarly changed in CCA tumors from two independent cohorts of patients, and in tumor cells and CCA-derived EVs in vitro, highlighting their utility as liquid biopsy biomarkers as well as their potential value as targets for therapy.

High-throughput omic approaches have been of great help in order to find potential new candidate biomarkers. In fact, the identification of the proteomic content of EVs, as well as certain circulating proteins, in biofluids have already provided candidate biomarkers in bile, serum, and urine from patients with CCA.[18,29–31] We have recently described the differential proteomic profiles of serum EVs from patients with CCA, compared to HCC, PSC, and healthy individuals, reporting new potential protein biomarkers with high diagnostic capacity that must be internationally validated by ELISA technology.[18] Nevertheless, a full transcriptomic analysis in distinct body fluids (in particular EVs) from these patients has never been conducted and might result in the identification of novel, accurate biomarkers for the diagnosis of CCA. Although proteins are usually more stable than mRNAs, their presence within EVs provides them protection from degradation; moreover, RNAs are usually easier to detect and quantify, even when found at very low levels, which may help in their faster translation into the clinic.[32] In fact, circulating small ncRNAs are found in all biofluids (including serum and urine), mainly due to their

remarkable resistance to RNase degradation. High circulating RNase levels contribute to a low abundance of other types of RNAs (mRNAs) and significantly compromise their easy and reliable detection.[33] Still, specific RNAs can be released from cancer cells into biofluids, allowing their identification and further determination of their potential value as biomarkers, as they may mirror the cellular state within the concept of liquid biopsy. For instance, some RNA transcripts were already evidenced and found increased in plasma, as is the case of telomerase reverse transcriptase (hTERT) that showed diagnostic and prognostic value for prostate cancer, being a good predictor of recurrence.[34,35] Similarly, the levels of the long non-coding RNA prostate cancer associated 3 (PCA3) were abundantly found in urine of patients with prostate cancer, constituting a promising non-invasive biomarker for the diagnosis of that cancer.[36,37] The levels of several miRNAs were also reported altered in serum, plasma, and urine of patients with gastrointestinal cancers, including CCA, constituting also potential novel biomarkers for cancer diagnosis.[7,32,38–40]

Taking advantage from our previously reported EV isolation protocol,[18] we have here also settled up the protocol for the isolation of urine EVs. By analyzing the transcriptomic profile of serum and urine EVs from patients with CCA, the present study was pioneer in identifying novel potential RNA biomarkers with high diagnostic capacity for CCA. Noteworthy, some of these new RNA biomarkers were also significantly altered in CCA tissue from the two international cohorts of patients and further disturbed in CCA cell lines and in EVs secreted from these tumor cell lines, when compared with NHCs. Consequently, these biomarkers are mirroring what is happening in the tumor tissue, since the disturbances observed in tumor biopsies (and CCA cell lines) that are later secreted in EVs and released into the bloodstream are amenable for detection either in serum or urine (**Figure 3.15**). This constitutes a novel and innovative liquid biopsy approach where we are able to detect specific alterations that are observed in tumor tissue without obtaining tumor samples, harboring a high diagnostic value. In the future, evaluating the relevance of these biomarkers in predicting prognosis and in guiding therapeutic decisions is envisioned.

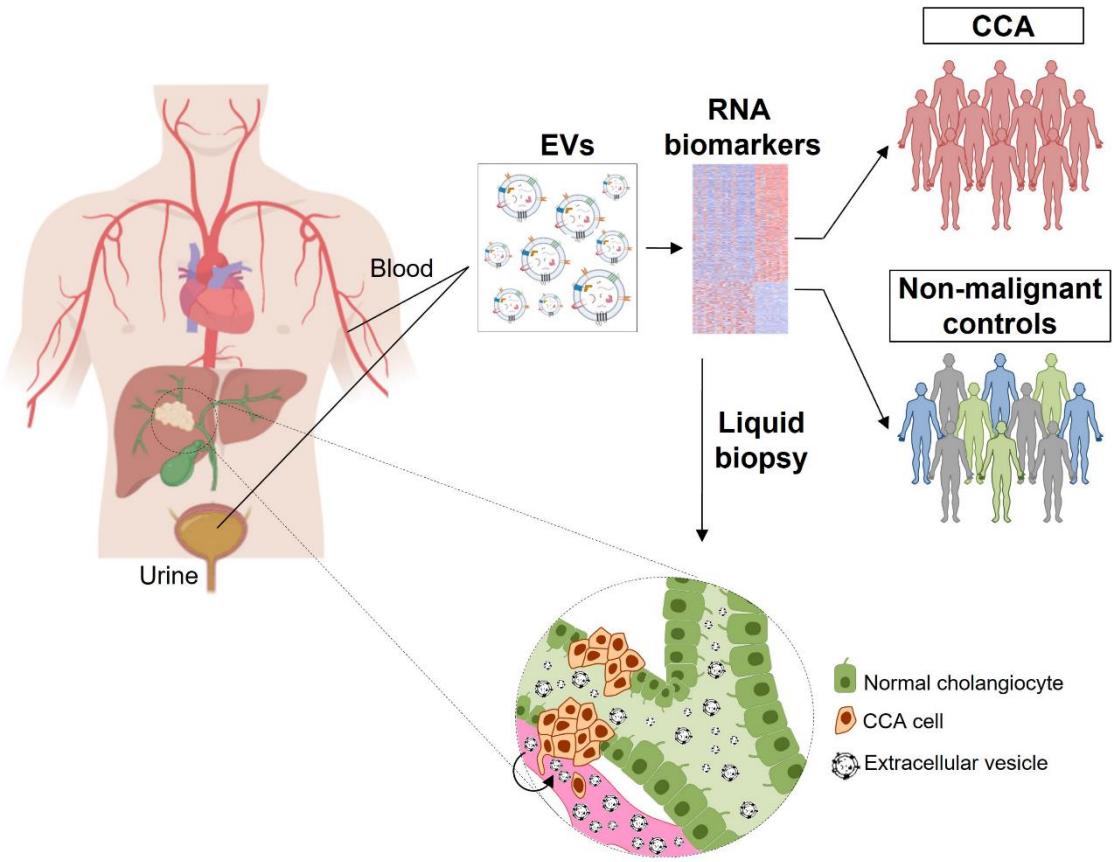


Figure 3.15. Novel liquid biopsy approach for cholangiocarcinoma. CCA tumor cells display distinct RNA expression profiles which are later released into circulation in EVs, containing potential biomarkers for CCA, that are amenable for detection in serum and urine, thus constituting a novel liquid biopsy approach.

Taking into consideration the 105 potential liquid biopsy biomarkers that were identified in serum, we herein reported the best five serum biomarkers that display excellent diagnostic accuracy: CMIP, GAD1, NME1, CDS1, and CKS1B. Importantly, these novel potential biomarkers might constitute better CCA biomarkers than CA19-9 since the AUC values that we herein obtained (up to 0.891) are higher than the diagnostic capacity reported in a systematic review and a meta-analysis, in which the AUC value for CA19-9 was 0.830.[41] Despite the potential diagnostic value of these transcripts, considering that they are concomitantly increased in tumor tissue in two international independent cohorts, we postulate that they also might play a pivotal pathological role during cholangiocarcinogenesis. Still, no studies have currently addressed the involvement of these biomolecules in CCA, although several works already valued their role in other types of cancer. For instance, the transcription factor CMIP was previously shown to be increased in human gastric cancer and glioma tumors, contributing to tumor proliferation and metastasis.[42,43] Furthermore, high CMIP levels were associated with worse prognosis (recurrence-free and overall survival) in gastric and breast cancers and were related with herceptin resistance in HER2-positive gastric cancer cells.[42,44,45] Similarly,

the enzyme encoded by GAD1 gene, which catalyzes the conversion of L-glutamic acid to γ-aminobutyric acid, has been found overexpressed in several types of tumors, including lung adenocarcinoma,[46] nasopharyngeal carcinoma,[47] oral squamous cell carcinoma,[48] prostate cancer,[49] and brain metastasis,[50] being also found upregulated in colon and HCC cells in vitro.[51] In parallel, high GAD1 levels were also shown to correlate with the pathological stage of patients with lung adenocarcinoma, positively correlating with metastasis and with worse recurrence-free survival.[46] Regarding NME1, its relevance in cancer is still controversial. A meta-analysis evaluated the prognostic value of NME1 in patients with digestive system neoplasms (including patients with HCC and gallbladder cancer, but not CCA) and reported that high NME1 levels are correlated with well-differentiated tumors and with less-severe cancer stages, with no evident correlation with prognosis.[52] In agreement with the reported metastasis-suppressing role of NME1, patients with HCC presenting lower protein levels of NME1 displayed increased metastatization.[53] In patients with HCC and in two animal models of HCC,[54,55] NME1 expression was found upregulated in comparison with non-tumor tissue, while knocking-out NME1 resulted in an increased number of lung metastasis.[55] Still, increased levels of NME1 negatively correlated with disease stage while NME1 upregulation was positively correlated with poor overall survival and with higher recurrence.[54] Furthermore, in melanoma patients, a NME1-related gene expression signature was reported and linked with increased overall survival,[56] pinpointing for a potential tumor suppressor role, but new pro-oncogenic roles, mainly related with the expansion of stem-like features and tumor growth, were recently reported.[57] Considering CDS1, a DNA damage-related function is well-known and a screening of several cancer cell lines highlighted that the expression levels of CDS1 are dependent on p53 activity, being inversely correlated with the presence of a functional p53.[58–60] Although CDS1 is thought to act as a tumor suppressor, there is no much information in this field. For instance, CDS1 was shown to be decreased in patients with HCC, likely as a result of promoter hypermethylation.[61] Nevertheless, its relevance for hepato- and/or cholangiocarcinogenesis is yet to be unveiled. Additionally, CKS1B represents an oncogene that has been reported increased in retinoblastoma, HCC, nasopharyngeal carcinoma and multiple myeloma.[62–66] Importantly, CKS1B upregulation has been positively related to increased proliferation, migration, angiogenesis, invasion and chemoresistance, being further related with lymph node metastasis and worse prognosis. Overall, regarding these genes, although nothing is still reported in CCA pathogenesis, they potentially constitute new key carcinogenic players that deserve further attention in the near future.

Considering the potential urine liquid biopsy biomarkers, we herein were able to report 39 transcripts that are differentially found in EVs from patients with CCA, sharing their

pattern of expression in CCA patients from the TCGA and Copenhagen cohorts and in CCA cells and in CCA-derived EVs in vitro. In fact, this number of transcripts corresponds to almost one-third of the identified serum liquid biopsy biomarkers, which might indicate that serum constitutes a richer source of potential CCA biomarkers. In this line, we have to bear in mind that EVs released from tumor masses directly enter into the bloodstream and their analysis might better mirror what is currently happening in the tumor. Furthermore, glomerular filtration might impact on the detected tumor-released EVs, therefore explaining the reduced number of urine liquid biopsy biomarkers. Still, urine analysis is considered the least invasive procedure for the diagnosis of cancer so far and we here described some potential new urine biomarkers with good diagnostic values that are also altered in tumor samples. For instance, UBE2C is commonly upregulated in intestinal-type gastric cancer, ovarian cancer, head and neck squamous cell carcinoma, pancreatic ductal adenocarcinoma, HCC, and non-small cell lung cancer.[67–73] Remarkably, increased UBE2C expression promotes chromosomal instability, cell cycle progression, proliferation and EMT, being positively correlated with worse clinical outcomes, mainly overall survival, lymph node metastasis, and progression-free survival. Of note, experimental inhibition of UBE2C suppressed the malignant phenotypes, surpassing cisplatin resistance in ovarian and non-small cell lung cancers,[68,73] in parallel with overcoming sorafenib resistance in HCC.[72] On the other hand, the reports regarding SERPINB1 are still controversial. SERPINB1 levels were found decreased in patients with prostate cancer, glioma, and HCC, favouring migration and invasion,[74–76] while displaying increased expression in cells and patients with oral cancer, correlating with high motility and cell migration in vitro.[77] Importantly, SERPINB1 levels were shown to predict the outcome of cisplatin-based chemotherapies in melanoma and the value of this transcript/protein as surrogate marker for CCA should be evaluated in the future.[78]

Besides identifying novel CCA biomarkers, we were also able to identify novel potential RNA transcripts for the differential diagnosis of CCA *vs*. PSC. Distinguishing between PSC-associated benign biliary strictures and early-stage CCA lesions is challenging. In fact, magnetic resonance imaging (MRI) has a limited resolution for this differential diagnosis and performing conventional cytology and/or fluorescence in situ hybridization (FISH) after invasive biliary brushing by endoscopic retrograde cholangiography only display a moderate diagnostic accuracy and increase the odds for procedure-related complications, including pancreatitis and cholangitis.[5,79] In fact, several serum and urine EV transcripts provided the maximum diagnostic capacity for the differential diagnosis of CCA and PSC, constituting potential biomarkers for the early diagnosis of CCA in patients with PSC. Although this work was mainly focused on the study of mRNA transcripts, several ncRNAs stood out as novel potential diagnostic biomarkers for the

diagnosis of CCA. In particular, the levels of the lncRNA MALAT1 are increased in serum EVs from patients with CCA compared to PSC, displaying an AUC value of 1.00. In this regard, MALAT1 was found upregulated in HCC, contributing to tumor development and progression,[80–83] as well as identified in HCC-derived exosomes.[84] MALAT1 is also increased in colon cancer cells, directly binding to miR-663a, thus acting as a competing endogenous lncRNA; as a result, important cancer-related miR-663a targets (TGFB1, PIK3CD, P21, JunB, and JunD) were affected.[85] These results encourage future studies on the potential role of MALAT1 in CCA. In addition, increased levels of miR-604 and miR-551B in serum EVs from patients with CCA also provided excellent diagnostic capacities (up to 0.944) for CCA and some studies already started to evaluate their involvement in carcinogenesis and as diagnostic biomarkers.[86–89] Considering the high resistance of small ncRNAs to RNase degradation, additional efforts should be employed in order to validate their diagnostic accuracy.

Overall, here we reported for the first time the differential RNA profiles of serum and urine EVs of patients with CCA, PSC, or UC, compared to healthy individuals, identifying novel non-invasive accurate biomarkers for CCA that mirror their expression in tumor tissue, thus constituting a novel liquid biopsy approach. The newly identified RNA biomarkers might markedly facilitate the diagnosis of these diseases and should be taken into consideration in future studies. These results pave the path not only for the discovery of new biomarkers but also for new potential therapeutic targets, as they may participate in disease pathogenesis. However, these results need to be now validated in large, international, and well-characterized cohorts of patients (and also including patients with CCA on a PSC background), in order to ascertain the diagnostic accuracy of the proposed biomarkers and then translate our findings into clinics. More importantly, it is pivotal to validate these results with easy transferable techniques [*e.g.*, quantitative PCR (qPCR) and droplet digital PCR (ddPCR), among others], in order to avoid the time-consuming steps and equipment demands for EVs isolation. In this way, we will be a little bit closer to see the use of accurate non-invasive diagnostic biomarkers for CCA at the clinics.

REFERENCES

1. Banales JM, Cardinale V, Carpino G, et al. Expert consensus document: Cholangiocarcinoma: current knowledge and future perspectives consensus statement from the European Network for the Study of Cholangiocarcinoma (ENS-CCA). *Nat Rev Gastroenterol Hepatol*. 2016;13(5):261-280. doi:10.1038/NRGASTRO.2016.51

2. Rizvi S, Khan SA, Hallemeier CL, Kelley RK, Gores GJ. Cholangiocarcinoma - evolving concepts and therapeutic strategies. *Nat Rev Clin Oncol*. 2018;15(2):95-111. doi:10.1038/NRCLINONC.2017.157

3. Bertuccio P, Malvezzi M, Carioli G, et al. Global trends in mortality from intrahepatic and extrahepatic cholangiocarcinoma. *J Hepatol*. 2019;71(1):104-114. doi:10.1016/J.JHEP.2019.03.013

4. Karlsen TH, Folseraas T, Thorburn D, Vesterhus M. Primary sclerosing cholangitis - a comprehensive review. *J Hepatol*. 2017;67(6):1298-1323. doi:10.1016/J.JHEP.2017.07.022

5. Razumilava N, Gores GJ. Surveillance for Cholangiocarcinoma in Patients with Primary Sclerosing Cholangitis: Effective and Justified? *Clin liver Dis*. 2016;8(2):43-47. doi:10.1002/CLD.567

6. Marin JJG, Lozano E, Herraez E, et al. Chemoresistance and chemosensitization in cholangiocarcinoma. *Biochim Biophys acta Mol basis Dis*. 2018;1864(4 Pt B):1444-1453. doi:10.1016/J.BBADIS.2017.06.005

7. Macias RIR, Kornek M, Rodrigues PM, et al. Diagnostic and prognostic biomarkers in cholangiocarcinoma. *Liver Int*. 2019;39 Suppl 1(S1):108-122. doi:10.1111/LIV.14090

8. Hirschfield GM, Karlsen TH, Lindor KD, Adams DH. Primary sclerosing cholangitis. *Lancet (London, England)*. 2013;382(9904):1587-1599. doi:10.1016/S0140-6736(13)60096-3

9. Lazaridis KN, LaRusso NF. Primary Sclerosing Cholangitis. Ingelfinger JR, ed. *N Engl J Med*. 2016;375(12):1161-1170. doi:10.1056/NEJMRA1506330

10. Macias RIR, Banales JM, Sangro B, et al. The search for novel diagnostic and prognostic biomarkers in cholangiocarcinoma. *Biochim Biophys acta Mol basis Dis*. 2018;1864(4 Pt B):1468-1477. doi:10.1016/J.BBADIS.2017.08.002

11. Eloubeidi MA, Chen VK, Jhala NC, et al. Endoscopic ultrasound-guided fine needle aspiration biopsy of suspected cholangiocarcinoma. *Clin Gastroenterol Hepatol*. 2004;2(3):209-213. doi:10.1016/S1542-3565(04)00005-9

12. Lapitz A, Arbelaiz A, Olaizola P, et al. Extracellular Vesicles in Hepatobiliary Malignancies. *Front Immunol*. 2018;9(OCT). doi:10.3389/FIMMU.2018.02270

13. Yáñez-Mó M, Siljander PRM, Andreu Z, et al. Biological properties of extracellular vesicles and their physiological functions. *J Extracell vesicles*. 2015;4(2015):1-60. doi:10.3402/JEV.V4.27066

14. Raposo G, Stoorvogel W. Extracellular vesicles: exosomes, microvesicles, and friends. *J Cell Biol*. 2013;200(4):373-383. doi:10.1083/JCB.201211138

15. González E, Falcón-Pérez JM. Cell-derived extracellular vesicles as a platform to identify low-invasive disease biomarkers. *Expert Rev Mol Diagn*. 2015;15(7):907-923. doi:10.1586/14737159.2015.1043272

16. Hirsova P, Ibrahim SH, Verma VK, et al. Extracellular vesicles in liver pathobiology: Small particles with big impact. *Hepatology*. 2016;64(6):2219-2233. doi:10.1002/HEP.28814

17. Masyuk AI, Huang BQ, Ward CJ, et al. Biliary exosomes influence cholangiocyte regulatory mechanisms and proliferation through interaction with primary cilia. *Am J Physiol Gastrointest Liver Physiol*. 2010;299(4). doi:10.1152/AJPGI.00093.2010

18. Arbelaiz A, Azkargorta M, Krawczyk M, et al. Serum extracellular vesicles contain protein biomarkers for primary sclerosing cholangitis and cholangiocarcinoma. *Hepatology*. 2017;66(4):1125-1143. doi:10.1002/HEP.29291

19. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: management of cholestatic liver diseases. *J Hepatol*. 2009;51(2):237-267. doi:10.1016/J.JHEP.2009.04.009

20. Farshidfar F, Zheng S, Gingras MC, et al. Integrative Genomic Analysis of Cholangiocarcinoma Identifies Distinct IDH-Mutant Molecular Profiles. *Cell Rep*. 2017;19(13):2878-2880. doi:10.1016/J.CELREP.2017.06.008

21. Andersen JB, Spee B, Blechacz BR, et al. Genomic and genetic characterization of cholangiocarcinoma identifies therapeutic targets for tyrosine kinase inhibitors. *Gastroenterology*. 2012;142(4). doi:10.1053/J.GASTRO.2011.12.005

22. O'Rourke CJ, Matter MS, Nepal C, et al. Identification of a Pan-Gamma-Secretase Inhibitor Response Signature for Notch-Driven Cholangiocarcinoma. *Hepatology*. 2020;71(1):196-213. doi:10.1002/HEP.30816

23. Urribarri AD, Munoz-Garrido P, Perugorria MJ, et al. Inhibition of metalloprotease hyperactivity in cystic cholangiocytes halts the development of polycystic liver diseases. *Gut*. 2014;63(10):1658-1667. doi:10.1136/GUTJNL-2013-305281

24. Banales JM, Sáez E, Úriz M, et al. Up-regulation of microRNA 506 leads to decreased Cl-/HCO3- anion exchanger 2 expression in biliary epithelium of patients with primary biliary cirrhosis. *Hepatology*. 2012;56(2):687-697. doi:10.1002/HEP.25691

25. Munoz-Garrido P, Marin JJG, Perugorria MJ, et al. Ursodeoxycholic acid inhibits hepatic cystogenesis in experimental models of polycystic liver disease. *J Hepatol*. 2015;63(4):952-961. doi:10.1016/J.JHEP.2015.05.023

26. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. 2008;24(13):1547-1548. doi:10.1093/BIOINFORMATICS/BTN224

27. Pathan M, Keerthikumar S, Ang CS, et al. FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics*. 2015;15(15):2597-2601. doi:10.1002/PMIC.201400515

28. Pathan M, Keerthikumar S, Chisanga D, et al. A novel community driven software for functional enrichment analysis of extracellular vesicles data. *J Extracell vesicles*. 2017;6(1). doi:10.1080/20013078.2017.1321455

29. Lankisch TO, Metzger J, Negm AA, et al. Bile proteomic profiles differentiate cholangiocarcinoma from primary sclerosing cholangitis and choledocholithiasis. *Hepatology*. 2011;53(3):875-884. doi:10.1002/HEP.24103

30. Li L, Masica D, Ishida M, et al. Human bile contains microRNA-laden extracellular vesicles that can be used for cholangiocarcinoma diagnosis. *Hepatology*. 2014;60(3):896-907. doi:10.1002/HEP.27050

31. Metzger J, Negm AA, Plentz RR, et al. Urine proteomic analysis differentiates cholangiocarcinoma from primary sclerosing cholangitis and other benign biliary disorders. *Gut*. 2013;62(1):122-130. doi:10.1136/GUTJNL-2012-302047

32. Xi X, Li T, Huang Y, et al. RNA Biomarkers: Frontier of Precision Medicine for Cancer.

Non-coding RNA. 2017;3(1). doi:10.3390/NCRNA3010009

33. Tsui NBY, Ng EKO, Lo YMD. Stability of endogenous and added RNA in blood specimens, serum, and plasma. *Clin Chem*. 2002;48(10):1647-1653.

34. March-Villalba JA, Martínez-Jabaloyas JM, Herrero MJ, Santamaria J, Aliño SF, Dasí F. Cell-free circulating plasma hTERT mRNA is a useful marker for prostate cancer diagnosis and is associated with poor prognosis tumor characteristics. *PLoS One*. 2012;7(8). doi:10.1371/JOURNAL.PONE.0043470

35. March-Villalba JA, Martínez-Jabaloyas JM, Herrero MJ, Santamaría J, Aliño SF, Dasí F. Plasma hTERT mRNA discriminates between clinically localized and locally advanced disease and is a predictor of recurrence in prostate cancer patients. *Expert Opin Biol Ther*. 2012;12 Suppl 1(SUPPL. 1). doi:10.1517/14712598.2012.685716

36. Bussemakers MJG, van Bokhoven A, Verhaegh GW, et al. DD3::A New Prostate-specific Gene, Highly Overexpressed in Prostate Cancer. *Cancer Res*. 1999;59(23).

37. Hessels D, Klein Gunnewiek JMT, Van Oort I, et al. DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. *Eur Urol*. 2003;44(1):8-16. doi:10.1016/S0302-2838(03)00201-X

38. Esparza-Baquer A, Labiano I, Bujanda L, Perugorria MJ, Banales JM. MicroRNAs in cholangiopathies: Potential diagnostic and therapeutic tools. *Clin Res Hepatol Gastroenterol*. 2016;40(1):15-27. doi:10.1016/J.CLINRE.2015.10.001

39. Olaizola P, Lee-Law PY, Arbelaiz A, et al. MicroRNAs and extracellular vesicles in cholangiopathies. *Biochim Biophys acta Mol basis Dis*. 2018;1864(4 Pt B):1293-1307. doi:10.1016/J.BBADIS.2017.06.026

40. Chauhan R, Lahiri N. Tissue- and Serum-Associated Biomarkers of Hepatocellular Carcinoma. *Biomark Cancer*. 2016;8(Suppl 1):BIC.S34413. doi:10.4137/BIC.S34413

41. Liang B, Zhong L, He Q, et al. Diagnostic Accuracy of Serum CA19-9 in Patients with Cholangiocarcinoma: A Systematic Review and Meta-Analysis. *Med Sci Monit*. 2015;21:3555-3563. doi:10.12659/MSM.895040

42. Zhang J, Huang J, Wang X, et al. CMIP is oncogenic in human gastric cancer cells. *Mol Med Rep*. 2017;16(5):7277-7286. doi:10.3892/MMR.2017.7541

43. Wang B, Wu ZS, Wu Q. CMIP Promotes Proliferation and Metastasis in Human Glioma. *Biomed Res Int*. 2017;2017. doi:10.1155/2017/5340160

44. Wang CCN, Li CY, Cai JH, et al. Identification of Prognostic Candidate Genes in Breast Cancer by Integrated Bioinformatic Analysis. *J Clin Med*. 2019;8(8). doi:10.3390/JCM8081160

45. Xiang R, Han X, Ding K, Wu Z. CMIP promotes Herceptin resistance of HER2 positive gastric cancer cells. *Pathol Res Pract*. 2020;216(2). doi:10.1016/J.PRP.2019.152776

46. Tsuboi M, Kondo K, Masuda K, et al. Prognostic significance of GAD1 overexpression in patients with resected lung adenocarcinoma. *Cancer Med*. 2019;8(9):4189-4199. doi:10.1002/CAM4.2345

47. Lee YY, Chao TB, Sheu MJ, et al. Glutamate Decarboxylase 1 Overexpression as a Poor Prognostic Factor in Patients with Nasopharyngeal Carcinoma. *J Cancer*. 2016;7(12):1716-1723. doi:10.7150/JCA.15667

48. Kimura R, Kasamatsu A, Koyama T, et al. Glutamate acid decarboxylase 1 promotes metastasis of human oral cancer by β-catenin translocation and MMP7 activation. *BMC*

Cancer. 2013;13. doi:10.1186/1471-2407-13-555

49. Jaraj SJ, Augsten M, Häggarth L, et al. GAD1 is a biomarker for benign and malignant prostatic tissue. *Scand J Urol Nephrol*. 2011;45(1):39-45. doi:10.3109/00365599.2010.521189

50. Schnepp PM, Lee DD, Guldner IH, et al. GAD1 Upregulation Programs Aggressive Features of Cancer Cell Metabolism in the Brain Metastatic Microenvironment. *Cancer Res*. 2017;77(11):2844-2856. doi:10.1158/0008-5472.CAN-16-2289

51. H Y, G T, H W, et al. DNA methylation reactivates GAD1 expression in cancer by preventing CTCF-mediated polycomb repressive complex 2 recruitment. *Oncogene*. 2016;35(30):4020. doi:10.1038/ONC.2016.28

52. Han W, Shi CT, Cao FY, et al. Prognostic Value of NME1 (NM23-H1) in Patients with Digestive System Neoplasms: A Systematic Review and Meta-Analysis. *PLoS One*. 2016;11(8). doi:10.1371/JOURNAL.PONE.0160547

53. Iizuka N, Oka M, Noma T, Nakazawa A, Hirose K, Suzuki T. NM23-H1 and NM23-H2 Messenger RNA Abundance in Human Hepatocellular Carcinoma. *Cancer Res*. 1995;55(3).

54. Yang J, Lv Z, Huang J, Zhao Y, Li Y. High expression of NME1 correlates with progression and poor prognosis in patients of hepatocellular carcinoma. *Int J Clin Exp Pathol*. 2017;10(8):8561. Accessed November 24, 2021. /pmc/articles/PMC6965413/

55. Boissan M, Wendum D, Arnaud-Dabernat S, et al. Increased lung metastasis in transgenic NM23-Null/SV40 mice with hepatocellular carcinoma. *J Natl Cancer Inst*. 2005;97(11):836-845. doi:10.1093/JNCI/DJI143

56. Leonard MK, McCorkle JR, Snyder DE, et al. Identification of a gene expression signature associated with the metastasis suppressor function of NME1: prognostic value in human melanoma. *Lab Invest*. 2018;98(3):327-338. doi:10.1038/LABINVEST.2017.108

57. Wang Y, Kathryn Leonard M, Snyder DE, Fisher ML, Eckert RL, Kaetzel DM. NME1 Drives Expansion of Melanoma Cells with Enhanced Tumor Growth and Metastatic Properties. *Mol Cancer Res*. 2019;17(8):1665-1674. doi:10.1158/1541-7786.MCR-18-0019

58. Lee JS, Collins KM, Brown AL, Lee CH, Chung JH. hCds1-mediated phosphorylation of BRCA1 regulates the DNA damage response. *Nature*. 2000;404(6774):201-204. doi:10.1038/35004614

59. Brown AL, Lee CH, Schwarz JK, Mitiku N, Piwnica-Worms H, Chung JH. A human Cds1-related kinase that functions downstream of ATM protein in the cellular response to DNA damage. *Proc Natl Acad Sci U S A*. 1999;96(7):3745-3750. doi:10.1073/PNAS.96.7.3745

60. Tominaga K, Morisaki H, Kaneko Y, et al. Role of human Cds1 (Chk2) kinase in DNA damage checkpoint and its regulation by p53. *J Biol Chem*. 1999;274(44):31463-31467. doi:10.1074/JBC.274.44.31463

61. YEH K-T, TANG K-P, CHEN Y-L, SU W-W, WANG Y-F, CHANG J-G. Methylation Inactivates Expression of CDP-diacylglycerol Synthase 1 (CDS1) in Hepatocellular Carcinoma. *Cancer Genomics Proteomics*. 2006;3(3-4):231-238. Accessed November 24, 2021. https://cgp.iiarjournals.org/content/3/3-4/231

62. Zeng Z, Gao ZL, Zhang ZP, et al. Downregulation of CKS1B restrains the proliferation, migration, invasion and angiogenesis of retinoblastoma cells through the MEK/ERK signaling pathway. *Int J Mol Med*. 2019;44(1):103-114. doi:10.3892/IJMM.2019.4183

63. Kang YS, Jeong EJ, Seok HJ, et al. Cks1 regulates human hepatocellular carcinoma cell

progression through osteopontin expression. *Biochem Biophys Res Commun*. 2019;508(1):275-281. doi:10.1016/J.BBRC.2018.11.070

64. Xu L, Fan S, Zhao J, et al. Increased expression of Cks1 protein is associated with lymph node metastasis and poor prognosis in nasopharyngeal carcinoma. *Diagn Pathol*. 2017;12(1). doi:10.1186/S13000-016-0589-9

65. Shi L, Wang S, Zangari M, et al. Over-expression of CKS1B activates both MEK/ERK and JAK/STAT3 signaling pathways and promotes myeloma cell drug-resistance. *Oncotarget*. 2010;1(1):22-33. doi:10.18632/ONCOTARGET.105

66. Huang CW, Lin CY, Huang HY, et al. CKS1B overexpression implicates clinical aggressiveness of hepatocellular carcinomas but not p27(Kip1) protein turnover: an independent prognosticator with potential p27 (Kip1)-independent oncogenic attributes? *Ann Surg Oncol*. 2010;17(3):907-922. doi:10.1245/S10434-009-0779-8

67. Zhang J, Liu X, Yu G, et al. UBE2C Is a Potential Biomarker of Intestinal-Type Gastric Cancer With Chromosomal Instability. *Front Pharmacol*. 2018;9(AUG). doi:10.3389/FPHAR.2018.00847

68. Li J, Zhi X, Shen X, et al. Depletion of UBE2C reduces ovarian cancer malignancy and reverses cisplatin resistance via downregulating CDK1. *Biochem Biophys Res Commun*. 2020;523(2):434-440. doi:10.1016/J.BBRC.2019.12.058

69. Jin Z, Zhao X, Cui L, et al. UBE2C promotes the progression of head and neck squamous cell carcinoma. *Biochem Biophys Res Commun*. 2020;523(2):389-397. doi:10.1016/J.BBRC.2019.12.064

70. Wang X, Yin L, Yang L, et al. Silencing ubiquitin-conjugating enzyme 2C inhibits proliferation and epithelial-mesenchymal transition in pancreatic ductal adenocarcinoma. *FEBS J*. 2019;286(24):4889-4909. doi:10.1111/FEBS.15134

71. Wei ZI, Liu YI, Qiao SH, et al. Identification of the potential therapeutic target gene UBE2C in human hepatocellular carcinoma: An investigation based on GEO and TCGA databases. *Oncol Lett*. 2019;17(6):5409-5418. doi:10.3892/OL.2019.10232

72. Xiong Y, Lu J, Fang Q, et al. UBE2C functions as a potential oncogene by enhancing cell proliferation, migration, invasion, and drug resistance in hepatocellular carcinoma cells. *Biosci Rep*. 2019;39(4). doi:10.1042/BSR20182384

73. Wu Y, Jin D, Wang X, et al. UBE2C Induces Cisplatin Resistance via ZEB1/2-Dependent Upregulation of ABCG2 and ERCC1 in NSCLC Cells. *J Oncol*. 2019;2019. doi:10.1155/2019/8607859

74. Lerman I, Ma X, Seger C, et al. Epigenetic Suppression of SERPINB1 Promotes Inflammation-Mediated Prostate Cancer Progression. *Mol Cancer Res*. 2019;17(4):845-859. doi:10.1158/1541-7786.MCR-18-0638

75. Huasong G, Zongmei D, Jianfeng H, et al. Serine protease inhibitor (SERPIN) B1 suppresses cell migration and invasion in glioma cells. *Brain Res*. 2015;1600:59-69. doi:10.1016/J.BRAINRES.2014.06.017

76. Cui X, Liu Y, Wan C, et al. Decreased expression of SERPINB1 correlates with tumor invasion and poor prognosis in hepatocellular carcinoma. *J Mol Histol*. 2014;45(1):59-68. doi:10.1007/S10735-013-9529-0

77. Tseng MY, Liu SY, Chen HR, et al. Serine protease inhibitor (SERPIN) B1 promotes oral cancer cell motility and is over-expressed in invasive oral squamous cell carcinoma. *Oral Oncol*. 2009;45(9):771-776. doi:10.1016/J.ORALONCOLOGY.2008.11.013

78. Willmes C, Kumar R, Becker JC, et al. SERPINB1 expression is predictive for sensitivity and outcome of cisplatin-based chemotherapy in melanoma. *Oncotarget*. 2016;7(9):10117-10132. doi:10.18632/ONCOTARGET.6956

79. Razumilava N, Gores GJ. Cholangiocarcinoma. *Lancet (London, England)*. 2014;383(9935):2168-2179. doi:10.1016/S0140-6736(13)61903-0

80. Hou ZH, Xu XW, Fu XY, Zhou L Du, Liu SP, Tan DM. Long non-coding RNA MALAT1 promotes angiogenesis and immunosuppressive properties of HCC cells by sponging miR-140. *Am J Physiol Cell Physiol*. 2020;318(3):C649-C663. doi:10.1152/AJPCELL.00510.2018

81. Sonohara F, Inokawa Y, Hayashi M, et al. Prognostic Value of Long Non-Coding RNA HULC and MALAT1 Following the Curative Resection of Hepatocellular Carcinoma. *Sci Rep*. 2017;7(1). doi:10.1038/S41598-017-16260-1

82. Hou Z, Xu X, Zhou L, et al. The long non-coding RNA MALAT1 promotes the migration and invasion of hepatocellular carcinoma by sponging miR-204 and releasing SIRT1. *Tumour Biol*. 2017;39(7):1-11. doi:10.1177/1010428317718135

83. Konishi H, Ichikawa D, Yamamoto Y, et al. Plasma level of metastasis-associated lung adenocarcinoma transcript 1 is associated with liver damage and predicts development of hepatocellular carcinoma. *Cancer Sci*. 2016;107(2):149-154. doi:10.1111/CAS.12854

84. Dai X, Chen C, Xue J, et al. Exosomal MALAT1 derived from hepatic cells is involved in the activation of hepatic stellate cells via miRNA-26b in fibrosis induced by arsenite. *Toxicol Lett*. 2019;316:73-84. doi:10.1016/J.TOXLET.2019.09.008

85. Tian W, Du Y, Ma Y, Gu L, Zhou J, Deng D. MALAT1-miR663a negative feedback loop in colon cancer cell functions through direct miRNA-lncRNA binding. *Cell Death Dis*. 2018;9(9). doi:10.1038/S41419-018-0925-Y

86. Cheong JY ou., Shin HD o., Cho SW o., Kim YJ u. Association of polymorphism in microRNA 604 with susceptibility to persistent hepatitis B virus infection and development of hepatocellular carcinoma. *J Korean Med Sci*. 2014;29(11):1523-1527. doi:10.3346/JKMS.2014.29.11.1523

87. Bai S yang, Ji R, Wei H, et al. Serum miR-551b-3p is a potential diagnostic biomarker for gastric cancer. *Turk J Gastroenterol*. 2019;30(5):415-419. doi:10.5152/TJG.2019.17875

88. Zhang Y, Yan L, Han W. Elevated Level of miR-551b-5p is Associated With Inflammation and Disease Progression in Patients With Severe Acute Pancreatitis. *Ther Apher Dial*. 2018;22(6):649-655. doi:10.1111/1744-9987.12720

89. Guangyuan S, Hongcheng Z, Chenlin C, et al. miR-551b regulates epithelial-mesenchymal transition and metastasis of gastric cancer by inhibiting ERBB4 expression. *Oncotarget*. 2017;8(28):45725-45735. doi:10.18632/ONCOTARGET.17392

PART IV

PROTEIN MOLECULES IN SERUM EXTRACELLULAR VESICLES

CHAPTER 4

CIRCULATING VESICLES HOLD ETIOLOGY-RELATED PROTEIN BIOMARKERS OF CCA RISK, EARLY DIAGNOSIS AND PROGNOSIS MIRRORING TUMOR CELLS

Ainhoa Lapitz,[1] Mikel Azkargorta,[2,3] Ekaterina Zhuravleva,[4] Marit M. Grimsrud,[5] Colm J. O'Rourke,[4] Ander Arbelaiz,[1] Adelaida La Casta,[1] Mette Vesterhus,[5,6] Piotr Milkiewicz,[7,8] Malgorzata Milkiewicz,[8] Raul Jimenez-Agüero,[1] Tania Pastor,[1] Rocio I.R. Macias,[3,9] Ioana Riano,[1] Laura Izquierdo-Sanchez,[1] Marcin Krawczyk,[10,11] Cesar Ibarra,[12] Javier Bustamante,[12] Felix Elortza,[2,3] Juan M. Falcon-Perez,[3,13,14] Maria J. Perugorria,[1,3] Jesper B. Andersen,[4] Luis Bujanda,[1,3] Tom H. Karlsen,[5] Trine Folseraas,[5,15] Pedro M. Rodrigues,[1,3,14*] Jesus M. Banales[1,3,14,16*]

[1]Department of Liver and Gastrointestinal Diseases, Biodonostia Health Research Institute – Donostia University Hospital, University of the Basque Country (UPV/EHU), 20014, San Sebastian, Spain.
[2]Proteomics Platform, CIC bioGUNE, ProteoRed-ISCIII, Bizkaia Science and Technology Park, 48160, Derio, Spain.
[3]National Institute for the Study of Liver and Gastrointestinal Diseases (CIBERehd), ISCIII, 28220, Madrid, Spain.
[4]Biotech Research and Innovation Centre, Department of Health and Medical Sciences, University of Copenhagen, 2200, Copenhagen, Denmark.
[5]Norwegian PSC Research Center, Department of Transplantation Medicine, Division of Surgery, Inflammatory Medicine and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway.
[6]Dept. of Clinical Science, University of Bergen, Bergen, Norway.
[7]Liver and Internal Medicine Unit, Department of General, Transplant and Liver Surgery, Medical University of Warsaw, 02-097, Warsaw, Poland.
[8]Department of Medical Biology, Pomeranian Medical University, 70-204, Szczecin, Poland.
[9]Experimental Hepatology and Drug Targeting (HEVEPHARM), University of Salamanca, Biomedical Research Institute of Salamanca (IBSAL), 37007, Salamanca, Spain
[10]Department of Medicine II, Saarland University Medical Centre, Saarland University, 66412, Homburg, Germany.
[11]Laboratory of Metabolic Liver Diseases, Centre for Preclinical Research, Department of General, Transplant and Liver Surgery, 02-091, Warsaw, Poland.
[12]Hospital of Cruces, 48903, Bilbao, Spain.
[13]Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Exosomes Laboratory, 48160, Derio, Spain.
[14]Ikerbasque, Basque Foundation for Science, 48013, Bilbao, Spain.
[15]Section of Gastroenterology, Department of Transplantation Medicine, Oslo University Hospital, Oslo, Norway.
[16]Department of Biochemistry and Genetics, School of Sciences, University of Navarra, Pamplona.

Under review in *Journal of Hepatology*

ABSTRACT

Background & Aims: Cholangiocarcinomas (CCAs), heterogeneous biliary tumors with dismal prognosis, lack accurate early-diagnostic methods, especially important for individuals at high-risk (*i.e.*, primary sclerosing cholangitis (PSC)). Here, we aimed to identify precise non-invasive CCA biomarkers.

Methods: Serum extracellular vesicles (EVs) from patients with: i) isolated PSC (n=39); ii) PSC without clinical evidences of malignancy at sampling who developed CCA overtime (PSC to CCA; n=10); iii) concomitant PSC-CCA (n=14); iv) CCAs from non-PSC etiology (n=26); and v) healthy individuals (n=41) were analyzed by mass-spectrometry. Diagnostic biomarkers of PSC-CCA, non-PSC CCA or CCAs regardless etiology (pan-CCAs) were defined, and their expression evaluated in human multi-organs and within CCA tumors at single-cell level. Prognostic EV-biomarkers for CCA were described.

Results: High-throughput proteomics identified candidate diagnostic biomarkers for PSC-CCA, non-PSC CCA or pan-CCA, independent to sex, age and CCA subtype. Machine learning logit modelling disclosed PLCH1/FGL1 algorithm with diagnostic value of AUC=0.903 and OR=27.8 for early-stage PSC-CCA vs isolated PSC, overpowering CA19-9 (AUC=0.608, OR=2.0). An algorithm combining SAMP/A1AT allowed the diagnosis of early-stage non-PSC CCAs compared to healthy individuals (AUC=0.863, OR=18.5). Noteworthy, the levels of 6 proteins (ALBU;FIBB;FLG1;IGHA1;TLN1;IMA8) showed predictive value for CCA development in patients with PSC before clinical evidences of malignancy. Multi-organ transcriptomic analysis revealed that serum EV-biomarkers were mostly expressed in hepatobiliary tissues and scRNA-seq analysis of CCA tumors indicated that some biomarkers –including PIGR,FGG,SERPINA1,FGL1– were mainly expressed in malignant cholangiocytes. Multivariable analysis revealed EV-prognostic biomarkers independent to clinical features, with FCN2/SDPR/FA9 panel being strongly associated to patients' survival.

Conclusions: Serum EVs contain etiology-specific protein biomarkers for the prediction, early diagnosis and prognosis estimation of CCA, representing a novel tumor cell-derived liquid biopsy for personalized medicine.

1. INTRODUCTION

Cholangiocarcinoma (CCA) includes a heterogeneous group of malignant tumors that can emerge at any location of the biliary system, displaying dismal prognosis. According to the anatomical origin, CCAs are classified as intrahepatic (iCCA), perihilar (pCCA) or distal (dCCA).[1–3] CCA is still considered a rare type of cancer, however, in the last decades, its global incidence and related mortality rates have been alarmingly increasing, currently representing the second most frequent primary liver cancer.[4,5] Furthermore, the silent growth of CCAs strongly impact their early detection, compromising patients' accessibility to potentially curative options (*i.e.*, tumor resection).[1,2]

CCAs are usually diagnosed by combining imaging methods and cytological/histological analysis of the tumor.[6] The cancer biomarker carbohydrate antigen 19-9 (CA19-9) is the only liquid-biopsy tool currently used in clinics to help in CCA diagnosis, but its diagnostic power is very low, specially at early CCA stages. The suboptimal accuracy of current non-invasive diagnostic approaches reflects the need for cytological/histological confirmation. Nevertheless, tumor biopsy or brushing is sometimes discouraged due to patients' fragility and advanced disease stages, risk of bleeding and peritoneal seeding, and specially in pCCA and dCCA cases the low amount of tissue collected may not be sufficient for cytological/histological confirmation.[7]

Most CCAs are considered sporadic and lack clear etiology, although some well-established conditions significantly increase the odds of CCA development, including the presence of choledochal cysts, biliary stones, cirrhosis, viruses or biliary diseases (*e.g.*, primary sclerosing cholangitis (PSC)).[2,8] In particular, PSC represents a chronic, cholestatic and immune-mediated liver disease of unknown etiology characterized by liver cell death, fibrosis and hepatic failure. PSC confers a substantial risk of CCA development (up to 20% lifetime risk), resulting in premature death.[1–3,9,10] Early diagnosis of CCA in patients with PSC by non-invasive methods is challenging, as there are overlapping radiological features between benign and malignant biliary strictures. All these evidences highlight the need of accurate non-invasive biomarkers for CCAs as a way to establish surveillance programs for its early detection in high-risk populations and also to provide a faster diagnosis of sporadic CCAs, ultimately decreasing cancer-related mortality.

In this sense, extracellular vesicles (EVs) –*i.e.*, nanometer lipid-bilayered spheres–, have arisen as promising source of biomarkers for human diseases.[11,12] These vesicles that are released from cells and found in biofluids, contain distinct types of biomolecules (*e.g.*,

proteins, nucleic acids, lipids and metabolites), participating in cell-to-cell communication and being useful tools in biomarker discovery.[13–17] In this study, we aimed to characterize the proteomic profile of serum EVs from individuals with PSC-associated CCA (PSC-CCA), CCA regardless of disease etiology, PSC that developed CCA during follow-up (PSC to CCA), isolated PSC and healthy subjects as a way to identify accurate biomarkers to predict CCA development, to early diagnose these tumors, as well as to estimate prognosis of patients with CCA.

2. MATERIALS AND METHODS

2.1 Study Population

Serum from individuals with: *i*) isolated PSC (n=18); *ii*) concomitant PSC and CCA (PSC-CCA; n=14); *iii*) PSC who developed CCA during follow-up (PSC to CCA; n=10); and *iv*) healthy subjects (n=19) were obtained from Oslo University Hospital Rikshospitalet (Oslo, Norway), Donostia University Hospital (San Sebastian, Spain) and Medical University Hospital of Warsaw (Warsaw, Poland). Samples from patients with isolated PSC were obtained after confirmation of no CCA development for a >5-years period after diagnosis. Serum from patients with PSC-CCA was obtained <2 months prior CCA diagnosis or when tumor development was already confirmed. Serum within the PSC to CCA group was from patients with PSC and no clinical evidences of tumor presence at sampling, but who developed CCA during follow-up (sampling: 5-26 months before CCA diagnosis). The study protocol was approved by the Regional Committee for Medical and Health Research Ethics South Eastern Norway (6.2008.1723), the Euskadi Drug Research Ethics Committee (CEIm-E: PI2019116), the Bioethical Committees of Medical University of Warsaw (KB/58/A/2016) and the Pomeranian Medical University (BN-001/43/06). Informed consent was obtained from all individuals.

This collection of samples, named as “Set 1”, was combined with a previously-published serum EV-proteomic dataset generated by our group (“Set 2”[18]), resulting in the following final groups: *i*) isolated PSC (n=39); *ii*) PSC to CCA (n=10); *iii*) PSC-CCA (n=14); *iv*) non-PSC CCAs (n=26) and *iv*) healthy individuals (n=41). Clinical and biochemical features of the study population were analyzed at the time of serum collection and are summarized in **Table 4.1**. As indicated in the European Association for the Study of the Liver (EASL) guidelines,[19] PSC diagnosis was based on standard clinical, biochemical, cholangiographic and histological criteria. For all patients with CCA, tumor development

was confirmed by histology/cytology. Tumor stage was determined based on the 7th edition of the American Joint Committee on Cancer (AJCC) classification.[20]

Table 4.1. Demographic and clinical features of the study cohort

		Healthy (n=41)		PSC (n=39)		PSC to CCA (n=10)	PSC-CCA (n=14)	CCA (n=26)
		Set 1 (n=19)	Set 2 (n=22)	Set 1 (n=18)	Set 2 (n=21)	Set 1	Set 1	Set2
GENERAL DEMOGRAPHICS								
Age	mean ± SD (years)	**60** ± 6	**61** ± 5	**35** ± 13	**30** ± 8	**46** ± 12	**53** ± 12	**65** ± 11
Sex	Male, n (%)	**5** (26.3)	**12** (54.5)	**15** (83.3)	**16** (76.2)	**9** (90.0)	**10** (71.4)	**17** (65.4)
CLINICAL PARAMETERS, n (%)								
Tumor stage [AJCC 7th edition]	I-II	-	-	-	-	**2** (20.0)	**5** (35.7)	**15** (57.7)
	III-IV	-	-	-	-	**8** (80.0)	**9** (64.3)	**11** (42.3)
CCA subtype	iCCA	-	-	-	-	**3** (30.0)	**3** (21.4)	**8** (30.8)
	pCCA	-	-	-	-	**6** (60.0)	**7** (50.0)	**4** (15.4)
	dCCA	-	-	-	-	**1** (10.0)	**4** (28.6)	**14** (53.8)
Cirrhosis	Yes	-	-	**0** (0.0)	**5** (23.8)	**4** (40.0)	**1** (7.1)	**0** (0.0)
	No	-	-	**15** (83.3)	**16** (76.2)	**6** (60.0)	**12** (85.7)	**26** (100.0)
	NA	-	-	**3** (16.7)	-	-	**1** (7.1)	-
IBD	No	-	-	**2** (11.1)	**3** (14.3)	**3** (30.0)	**0** (0.0)	**26** (100.0)
	UC	-	-	**10** (55.6)	**17** (81.0)	**6** (60.0)	**13** (92.9)	**0** (0.0)
	Crohn	-	-	**3** (16.7)	**0** (0.0)	**1** (10.0)	**0** (0.0)	**0** (0.0)
	Unspecified	-	-	**3** (16.7)	**1** (4.8)	**0** (0.0)	**1** (7.1)	**0** (0.0)
Surgery	No	-	-	-	-	**2** (20.0)	**5** (35.7)	**2** (7.7)
	Liver transplant	-	-	-	-	**7** (70.0)	**6** (42.9)	**0** (0.0)
	Tumor resection	-	-	-	-	**1** (10.0)	**3** (21.4)	**18** (69.2)
	Exploratory	-	-	-	-	**0** (0.0)	**0** (0.0)	**1** (3.8)
	NA	-	-	-	-	-	-	**5** (19.2)
BIOCHEMICAL PARAMETERS								
ALT (IU/L)	mean ± SD	-	-	**108** ± 85.5	**124** ± 88.7	**108** ± 69.6	**129** ± 209	**106** ± 114
	NA, n (%)	-	-	-	-	-	**1** (7.1)	-
AST (IU/L)	mean ± SD	-	-	**69.8** ± 69.3	**97.2** ± 74.8	**95.3** ± 45.1	**142** ± 209	**67.4** ± 57.0
	NA, n (%)	-	-	-	-	-	-	-
GGT (IU/L)	mean ± SD	-	-	**301** ± 234	**308** ± 313	**305** ± 260	**250** ± 249	**711** ± 713
	NA, n (%)	-	-	-	-	-	**1** (7.1)	**2** (7.7)
Total bilirubin (mg/dL)	mean ± SD	-	-	**16.7** ± 16.1	**3.1** ± 5.7	**51.3** ± 92.6	**159.5** ± 161.2	**12.5** ± 27.6
	NA, n (%)	-	-	-	-	-	-	-
ALP (IU/L)	mean ± SD	-	-	**238** ± 161	**357** ± 293	**422** ± 253	**381** ± 340	**345** ± 255
	NA, n (%)	-	-	-	-	-	-	**4** (15.4)
CA19-9 (IU/mL)	mean ± SD	-	-	**31.5** ± 34.6	**433** ± 1731	**184** ± 451	**1815** ± 3152	**1530** ± 3709
	NA, n (%)	-	-	**3** (16.7)	-	-	**1** (7.1)	**2** (14.3)

2.2 Isolation and characterization of EVs from human serum samples

EV isolation and characterization procedures were guided by the International Society for Extracellular Vesicles statement[21] and carried out as we previously described, by serial differential ultracentrifugation.[18] Briefly, 1 mL of serum was diluted in phosphate buffered saline (PBS) (DPBS, Gibco) in ultracentrifugation tubes (thick wall polycarbonate centrifuge tubes, Beckman coulter), which were centrifuged in a TLA110

rotor (Beckman coulter) at 10,000xg for 30 min to remove cell debris and large EVs. The supernatant was then subsequently ultracentrifuged at 100,000xg for 75 min. Pelleted EV fraction was then washed with PBS and ultracentrifuged at 100,000xg for another 75 min. Finally, the pelleted EV fraction was resuspended in 20 µL of PBS and then stored at -80ºC for further analysis. All relevant data on isolation and characterization techniques have been submitted to the EV-TRACK knowledgebase (EV-TRACK ID: EV210077).[22]

2.2.1 Transmission Electron Microscopy (TEM)

For the characterization of EVs, PBS-resuspended EV isolate was negatively stained and evaluated by TEM. EV samples were directly adsorbed onto a glow-discharged (60 seg low discharging using a PELCO easy-glow device) carbon-coated copper grid (300 mesh). Afterwards, grids were fixed with 2% paraformaldehyde (PFA) in PBS 0.2M pH 7.4 for 20 min and washed with distilled water. Then, contrast staining was made by incubating the grids with 4% uranyl acetate at 4ºC for 15 min. TEM images were obtained by using TECNAI G2 20 C-TWIN high-resolution transmission electron microscope, at an acceleration voltage of 200 kV.

2.2.2 Nanoparticle tracking analysis (NTA)

To evaluate the size distribution and particle concentration of the isolate, EV samples were diluted 250-fold in PBS and later measured by using a NanoSight LM10 System (Malvern, UK) equipped with fast video capture and analyzed with the Nanoparticle Tracking Analysis (NTA) Version 2.2 particle-tracking software. Each EV preparation was measured twice and NTA post-acquisition settings were kept constant for all samples. Each recorded 1 min video was analyzed to obtain the mean and mode vesicle size as well as particle concentration.

2.2.3 Protein quantification, electrophoresis and immunoblotting

Protein concentration of isolated EVs, total serum, EV-depleted serum, as well as of whole cell extract (WCE) from normal human cholangiocyte (NHC) cultures was measured using the Micro BCA™ Protein Assay Kit according to the manufacturer's instructions (Thermo Scientific, USA).

In order to evaluate the expression of characteristic EV markers tetraspanins CD63 and CD81, 20 µg of serum EV fraction, total serum, EV-depleted serum and WCE samples

were denaturalized by adding non-reducing Protein Loading Buffer [50 mM Tris-HCl, 2% SDS, 10% glycerol and 0.1% bromophenol blue, without β-mercaptoethanol or dithiothreitol (DTT)] and by heating the samples at 95ºC for 5 min. For the detection of the negative control endoplasmic reticulum chaperone BiP (GRP78), 10 µg of protein was used and 500 mM of 2-mercaptoetahnol (Sigma-Aldrich) were included in the loading buffer. Then, proteins were separated in a 12.5% sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and electro-transferred onto a 0.45 µm pore-size nitrocellulose membrane (GE Healthcare). After blocking with 5% skim milk powder/Tris-buffered saline with 0.1% Tween 20 (TBS-T) (Sigma Aldrich) for 1 h at room temperature, membranes were incubated overnight at 4ºC with the appropriate primary antibodies [anti-CD81 (#555675, BD Biosciences), anti-CD63 (#H5C6, DSHB) and anti-GRP78 (#610979, BD Biosciences) at 1:500 dilution in blocking solution. Membranes were then washed 3 times with TBS-T and incubated with an anti-mouse horseradish peroxidase-conjugated secondary antibody (Cell Signalling, Danvers, MA, USA) at a dilution of 1:5000 (in blocking solution) for 1 h at room temperature. Membranes were washed with TBS-T and the signal was detected using the SuperSignal™ West Dura Extended Duration Substrate (Thermo Scientific, USA). Finally, the emitted chemiluminescence was visualized and captured with the iBright FL1500 Western Blot Imaging System (Thermo Fisher Scientific).

2.3 High-throughput proteomic analysis of serum EVs

2.3.1 Protein extraction and filter aided sample preparation (FASP)

All samples were resuspended in “cell lysis buffer” containing 7M Urea 2M Thiourea 4% CHAPS and incubated for 30 min under agitation. Next, FASP of the samples was performed mainly as previously described.[23] This protocol is based on the use of standard filtration devices allowing buffer exchange and acting as a reactor for the digestion. Each digestion step was followed by 20 min centrifugations at 13,000 rpm in order to remove the buffer from the filter. Samples were loaded onto Amicon Ultra 0.5 mL 30K centrifugal units (Millipore), washed twice in UA solution (8 M Urea, 100 mM Tris-HCl pH 8.5), reduced (20 min incubation in 100 mM DTT prepared in UA solution) and alkylated (20 min incubation in 50 mM Iodoacetamide prepared in UA solution). Then, 3 additional washes were carried out in UA, followed by 3 additional washes in 50 mM AMBIC. Protein was quantified using Bio-Rad protein assay (Bio-Rad), and trypsin was added to a trypsin:protein ratio of 1:10. The mixture was incubated overnight at

37ºC. Peptides were recovered from the filter units. Samples were speed-vacuumed in a RVC2 25 speed-vac concentrator (Christ). Zip-tip peptides were resuspended in 0.1% formic acid (FA) prior to mass spectrometry (MS) analysis.

2.3.2 Mass spectrometry analysis

Peptide separation was performed on a nanoACQUITY UPLC System (Waters) connected to a LTQ Orbitrap XL (Thermo Electron). An aliquot of each sample was loaded onto a Symmetry 300 C18 UPLC Trap column (180 µm x 20 mm, 5 µm) (Waters). The precolumn was connected to a BEH130 C18 column, 75 µm x 200 mm, 1.7 µm (Waters), and equilibrated in 3% acetonitrile and 0.1% FA. Peptides were eluted directly into the nanoelectrospray capillary (Proxeon Biosystems) at 300 nL/min, using a 120 min linear gradient of 3–50% acetonitrile.

The LTQ Orbitrap XL ETD automatically switched between MS and MS/MS acquisition in DDA mode. Full MS scan survey spectra (m/z 400–2000) were acquired in the orbitrap with mass resolution of 30000 at m/z 400. After each survey scan, the six most intense ions above 1000 counts were sequentially subjected to collision-induced dissociation (CID) in the linear ion trap. Precursors with charge states of 2 and 3 were specifically selected for CID. Peptides were excluded from further analysis for 60 seconds using the dynamic exclusion feature.

2.3.3 Progenesis liquid chromatography-MS (LC-MS) software analysis

Progenesis LC-MS (version 4.0.4265.42984, Nonlinear Dynamics) was used for the label-free differential protein expression analysis. Once the Raw files were imported, one of the runs was used as the reference to which the precursor masses in all other samples were aligned to. Only features comprising charges of 2+ and 3+ were selected. The raw abundances of each feature were automatically normalized and logarithmized against the reference run. A peak list containing the information of the detected different features was generated and exported to the Mascot search engine (Matrix Science Ltd.). The generated “mgf” file was searched against the Uniprot/Swissprot human database using 10 ppm and 0.5 Da tolerances for precursor and fragment ions, respectively. The list of identified peptides was imported in Progenesis LC-MS and the previously quantified features were matched to the corresponding peptides.

2.3.4 Peptide quantity normalization and batch effect correction by proBatch package in R

The relative quantification of non-conflicting peptides exported from Progenesis LC-MS was analyzed with the proBatch package in R in order to normalize peptide quantification data and to diagnose and correct for batch effects when needed (Set 1 + Set 2 meta-analysis).[24] Briefly, raw data matrix was transformed onto log scale and quality was examined by visualizing the global quantitative pattern plotting the sample mean and boxplots to identify discrepancies between batches. Next, in order to compared samples between set 1 and 2, the total intensity of the samples was scaled, normalizing the distribution of the raw intensities to be the same in all samples by applying quantile normalization. Then, as bias in the data can persist even after normalization, diagnostic plots for batch effects [*i.e.*, hierarchical clustering and principal component analysis (PCA) plots] were performed to evaluate to what extent technical variance still existed in the normalized data matrix. Next, the correction method based on parametric and non-parametric empirical Bayes framework ComBat was used for the batch effect adjustment procedure to make all samples comparable across batches. Finally, the correction of technical bias was evaluated by a quality control analysis on batch-corrected data matrix. For this evaluation, an exploratory correlation matrix was carried out with technical and biological replicates which were run in both sets.

2.3.5 Protein quantification of batch-corrected peptide data

As the correction procedure may alter the abundances of peptides that are critical for protein quantity inference, protein quantification was performed after batch effect correction of the peptide data. For this approach, protein quantification was calculated by the average of the antilogarithm corrected value of the peptides, which are unique for each protein. Finally, protein abundance was transformed into $\log_2$ scale and re-checked to confirm the batch effect correction at protein level by using the package proBatch as explained above.

2.4 Diagnostic biomarker selection

To assess the potential usefulness of serum EV proteins for the diagnosis of CCA in patients with PSC etiology or non-PSC etiologies, area under the receiver-operating characteristic curve (AUC) values were calculated for each biomolecule using SPSS

software version 22.0 (IBM, Ehningen, Germany). To classify candidate protein biomarkers into specific diagnostic biomarkers for CCA in patients with PSC (PSC-CCA diagnostic biomarkers), diagnostic biomarkers for CCA regardless its etiology (diagnostic biomarkers for ALL CCAs: pan-CCA) and diagnostic biomarkers for CCA of non-PSC etiologies (non-PSC CCA diagnostic biomarkers) the following criteria were established: a) if the levels of serum EV proteins correlated with the presence/absence of PSC, those proteins were analyzed as potential PSC-CCA or non-PSC CCA biomarkers b) biomarkers specific to PSC-CCA were selected from pair-wise comparisons of PSC-CCA to Healthy, PSC and non-PSC CCA groups, as well as PSC-CCA samples vs the rest based on AUC p-values; c) biomarkers specific to non-PSC CCA were selected from pair-wise comparisons of non-PSC CCA to Healthy, PSC and PSC-CCA groups, as well as non-PSC CCA samples vs the rest based on AUC p-values; d) if the levels of serum EV proteins did not correlate with the presence/absence of PSC, those proteins were analyzed as potential Pan-CCA biomarkers; e) biomarkers for the general diagnosis of CCA regardless its etiology (pan-CCA) were selected from pair-wise comparisons of pan-CCA to Healthy and PSC groups, as well as pan-CCA samples vs the rest based on AUC p-values.

The diagnostic parameters including sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predicting value (NPV), accuracy index (AI) and odds ratio (OR) were calculated by 2X2 contingency tables to patient groups dichotomized using the cut-off value established by the top Youden Index (YI) of each ROC curve. Additionally, multivariate binary logistic regression was accomplished to make logistic model protein biomarker combinations and to evaluate the diagnostic power of the logistic functions.

2.5 Gene expression analysis of circulating diagnostic biomarkers in human tissues

To evaluate the potential origin of serum EV protein biomarkers, the expression of those protein-coding genes was analyzed in human tissue bulk RNA datasets, in healthy liver single-cell RNA-seq (scRNA-seq) dataset, as well as in scRNA-seq data from patients with CCA.

2.5.1 Human multi-organ gene expression analysis of candidate biomarkers

Three transcriptomics datasets (HPA, GTEx and FANTOM5) downloaded from the Human Protein Atlas were used to estimate the relative gene expression levels of candidate biomarkers and assess the organ/tissue origin of the defined protein biomarkers derived from the serum EVs[25,26] (http://www.proteinatlas.org / v20.proteinatlas.org). The Human Protein Atlas (HPA) consortium RNA-seq dataset consists in mRNA sequencing of different 37 human tissues and 6 hematological cell types including B-cells, T-cells, NK-cells, monocytes, granulocytes and dendritic cells.[26] The Genotype-Tissue Expression (GTEx) project gathers RNA-seq data from 34 different human tissue and organs.[27] The Functional Annotation of Mammalian Genomes 5 (FANTOM5) project brings together mammalian cell-type specific transcriptomes from 45 different human tissues and organs by using the Cap Analysis of Gene Expression (CAGE) strategy.[28] RNA Consensus tissue gene data summarizes normalized transcript expression levels of 61 tissues based on transcriptomic data from the previous mentioned three sources: HPA, GTEx and FANTOM5. The normalized RNA Consensus data includes the following human tissue and organs: adipose tissue, adrenal gland, amygdala, appendix, B-cells, basal ganglia, bone marrow, breast, cerebellum, cerebral cortex, cervix, colon, corpus callosum, dendritic cells, ductus deferens, duodenum, endometrium, epididymis, esophagus, fallopian tube, gallbladder, granulocytes, heart muscle, hippocampal formation, hypothalamus, kidney, liver, lung, lymph node, midbrain, monocytes, natural killer (NK)-cells, olfactory region, ovary, pancreas, parathyroid gland, pituitary gland, placenta, pons and medulla, prostate, rectum, retina, salivary gland, seminal vesicle, skeletal muscle, skin, small intestine, smooth muscle, spinal cord, spleen, stomach, T-cells, testis, thalamus, thymus, thyroid gland, tongue, tonsil, total peripheral blood mononuclear cells (PBMCs), urinary bladder and vagina.

2.5.2 Single-cell RNA sequencing analysis (scRNA-seq)

Single-cell transcriptome profiling from healthy liver samples as well as from CCA tumors from two different studies were downloaded from GSE115469, GSE151530 and GSE125449, respectively.

The normal liver dataset (GSE115469) comprises the transcriptional profile of 8,444 cells obtained from the fractionation of fresh hepatic tissue from five healthy human livers.[29] From the scRNA-seq data published in GSE125449, a total number of 5,376 cells were chosen that were originally isolated from 10 tumor biopsies from patients with iCCA.[30]

Sequencing data from 4,961 cells coming from 14 fresh liver tumor biopsies from 12 patients with iCCA were also analyzed in the GSE151530 dataset.[31]

Regarding data processing, for the normal liver single-cell sequencing (GSE115469), data filtering and clustering was preserved as in the original study. In relation to CCA datasets, all these three datasets were processed with Seurat version 4.0.0. Briefly, features reported at least in 3 cells and cells with at least 500 features were considered. Cells with UMI counts below 700 or mitochondrial content above 35% in GSE 1515130 or above 20% in GSE125449 were removed. Outliers were defined per sample with scatter (more than 3 median absolute deviations from the median value) and removed. Doublets/multiplets were predicted and removed with scDblFinder, default settings. Data was scaled with LogNormalize transformation and scale factor 10,000 according to default Seurat settings. Variable genes for PCA were identified using the Seurat function FindVariableFeatures using the vst method with 2000 features selected in GSE 1515130 or with 2244 features selected in GSE125449. Data was scaled, UMI number and mitochondrial gene content were regressed out. For the tSNE plots, top 30 PCA components were selected to perform dimension reduction (dims.use = 1:30, resolution = 1) in GSE1515130 and top 20 in GSE125449 (dims.use = 1:20, resolution = 0.8).

Known cell marker genes used for classification of the main cell types were unchanged among all scRNA-seq studies: hepatocytes (*HPX*, *LBP*, *HPR* and *SERPINA10*); cholangiocytes (*KRT19*, *KRT7* and *FXYD2*); endothelial cells (*CD34*, *JAM2* and *VWF*); hepatic stellate cells (*COL1A2*, *LUM* and *DCN*); macrophages (*CD68*, *CD163* and *MSR1*); B cells (*CD19*, *FCRL5* and *CD79A*); NK cells (*KLRF1* and *NCAM1*) and T cells (*CD3D*, *CD3E* and *CD3G*) (**Figure 4.1**).

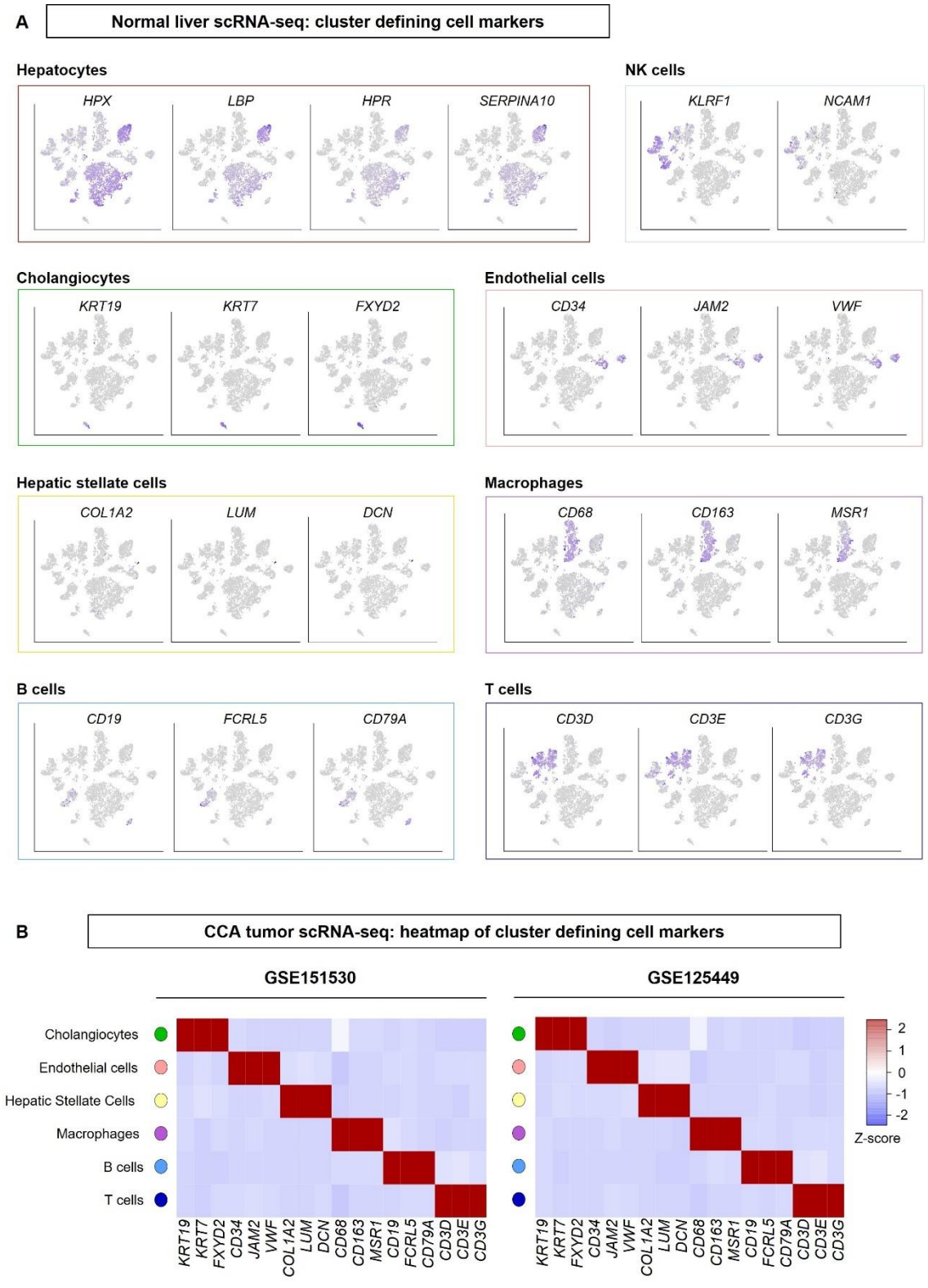


Figure 4.1. Expression of cluster-defining cell markers. (**A**) tSNE plots of cell type markers in normal liver scRNA-seq data. (**B**) Heatmaps of cell type markers in 2 CCA tumor scRNA-seq cohorts.

2.6 Survival analysis

The potential association of serum EV protein abundance with patients' OS was evaluated by univariable and multivariable Cox proportional hazard analysis and by the log rank (Mantel-Cox) test. For the Cox proportional hazard regression analysis, the R packages survival 3.2-7 and RegParallel 1.6.0 were used and unadjusted p-value <0.05 of the Hazard Ratio was considered as cut-off.

For the categorization of the continuous variables (*i.e.*, EV protein abundance), the Survminer 0.4.8 R package was applied. This package determines the optimal cut-point using the maximally selected rank statistics from the 'maxstat' R package.[32] The categorical classification of our study cohort according to high or low abundance of selected EV proteins in serum was employed to evaluate if protein dichotomic abundance was associated with OS time by performing the log rank (Mantel-Cox) statistical test comparing survival Kaplan-Meier curves as well as by calculating the Hazard Ratio of the Cox proportional-hazards model.

2.7 Statistical analysis

Statistical analyses were performed using R Studio with the R version 4.0.2 (2020-06-22), IBM SPSS Statistics software version 22 (IBM, Ehningen, Germany) and GraphPad Prism version 8.0 (GraphPad Software, San Diego, CA, USA). All values were tested to evaluate if they follow a Gaussian distribution by the normality tests D'Agostino-Pearson Omnibus and Shapiro-Wilk normality tests. When comparing two groups, non-parametric Mann-Whitney or parametric t-Student tests were conducted. For comparisons between more than two groups, non-parametric Kruskal-Wallis test followed by a posteriori Dunns test or the parametric one-way analysis of variance (ANOVA) test followed by a posteriori Tukey's post hoc test were used. Differences were considered significant when $p < 0.05$.

2.8 Data availability

All relevant data regarding EV isolation and characterization techniques have been submitted to the EV-TRACK knowledgebase (EV-TRACK ID: EV210077). Mass spectrometry data of serum EV proteomic data has been submitted to the database PRIDE – Proteomics Identification Database – EMBL-EBI (Project accession: "set 2" PXD026197; "set 1" PXD026199).

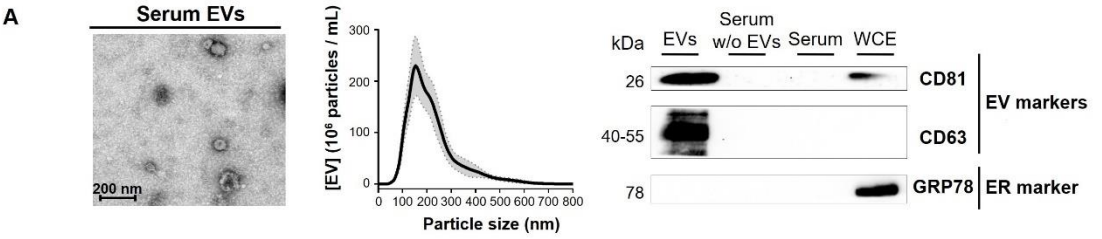
3. RESULTS

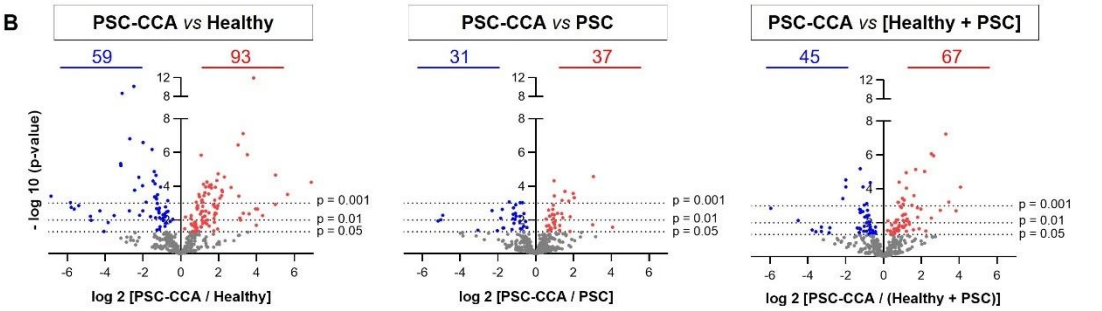
3.1 Differential abundance of EV-proteins from patients with PSC-CCA, PSC to CCA, PSC and healthy individuals revealed candidate biomarkers to predict and early diagnose CCA

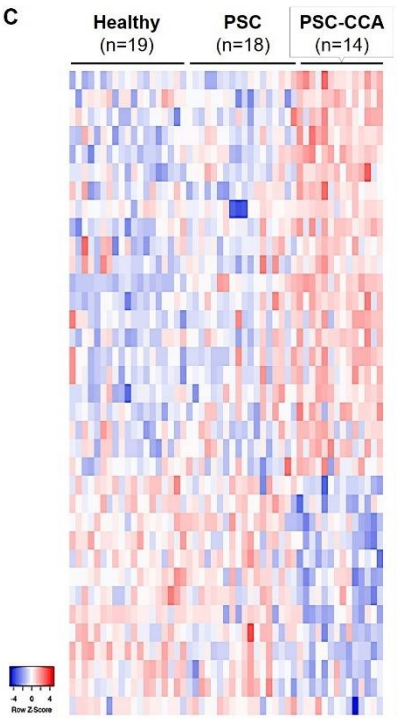
Serum EVs obtained from individuals with PSC-CCA, PSC to CCA, isolated PSC or healthy were characterized by transmission electron microscopy (TEM), nanoparticle tracking analysis (NTA) and immunoblotting (**Figure 4.2A**). TEM images confirmed the prevalence of classic cup-shaped rounded morphology vesicles with an average size smaller than 200 nm. By immunoblotting, EV-protein markers tetraspanins CD63 and CD81 were highly enriched in the isolated EV fraction when compared to total serum, EV-depleted serum (serum w/o EVs) and to whole-cell extract (WCE) from normal human cholangiocytes (NHCs), while the endoplasmic reticulum marker GRP78 was completely absent in isolated EVs, substantiating a proper isolation and high purity of the obtained EV fraction. Particle number measured by the light scattering technique NTA showed similar size of vesicles in all groups, with an average mean size of ~220 nm, and being particularly enriched in 170 nm particles (mode size), revealing exosomes and/or small and medium-size microvesicles as the principal components of the isolated EV fraction (**Figure 4.2A; Figure 4.3A**).

The proteomic profile of isolated EVs was then characterized by mass spectrometry. Univariable analysis revealed distinct EV-protein profiles when comparing individuals with PSC-CCA, isolated PSC or healthy (**Figure 4.2B-C**). Compared to healthy, the abundance of 152 proteins was altered in PSC-CCA (93 up; 59 down). Furthermore, the levels of 37 proteins were higher and 31 lower in PSC-CCA compared to isolated PSC. When comparing patients with PSC-CCA with a group composed by both healthy individuals and patients with PSC (“non-malignant control group”), the abundance of 67 proteins was found increased and 45 decreased in PSC-CCA (**Figure 4.2B**). Next, the diagnostic capacity of single candidate biomarkers was investigated. Several EV-proteins showed high AUC values for CCA diagnosis, with FRIL providing the highest score for PSC-CCA when compared to isolated PSC (0.909) or to the “non-malignant control group” (0.931) (**Figure 4.2C**). Of note, the abundance of several of these candidate diagnostic biomarkers, including FGL1, IGHA1, FRIL, FIBB, LAC2, FIBG, FIBA, PIGR and HEMO, were already found altered in EVs from patients with PSC who developed CCA during follow-up, when compared to patients with PSC who did not developed CCA, providing differential AUC values up to 0.900 and positive predictive

values (PPV) and negative predictive values (NPV) up to 77.8% and 100%, respectively (**Figure 4.2D; Figure 4.3C**).







Protein	PSC-CCA *vs* PSC			PSC-CCA *vs* Healthy			PSC-CCA *vs* [Healthy + PSC]		
	AUC	SEN (%)	SPE (%)	AUC	SEN (%)	SPE (%)	AUC	SEN (%)	SPE (%)
FRIL	0.909	85.7	88.9	0.951	85.7	94.7	0.931	85.7	91.9
LAC2	0.905	85.7	83.3	0.759	57.1	89.5	0.830	57.1	94.6
FGL1	0.869	85.7	88.9	0.895	78.6	94.7	0.882	78.6	91.9
LV301	0.853	85.7	83.3	0.904	92.9	79.0	0.879	85.7	83.8
OIT3	0.849	64.3	94.4	0.951	92.9	84.2	0.902	78.6	83.8
HV308	0.841	85.7	77.8	0.906	92.9	73.7	0.875	85.7	75.7
SVEP1	0.825	100.0	50.0	0.763	78.6	73.7	0.793	92.9	56.8
LFG3	0.796	78.6	83.3	0.782	78.6	84.2	0.789	78.6	83.8
IGHA1	0.790	64.3	88.9	0.876	85.7	84.2	0.834	85.7	75.7
K22E	0.790	92.9	66.7	0.722	92.9	63.2	0.755	92.9	72.7
LV403	0.786	71.4	77.8	0.733	92.9	57.9	0.759	92.9	56.8
PIGR	0.782	85.7	77.8	0.992	100.0	94.7	0.89	85.7	89.2
FCGBP	0.782	85.7	72.2	0.951	85.7	94.7	0.869	85.7	83.8
FIBG	0.778	85.7	77.8	0.797	85.7	79.1	0.788	85.7	78.4
IGLL5	0.770	78.6	72.2	0.823	64.3	89.5	0.797	64.3	83.8
FIBB	0.758	85.7	72.2	0.876	85.7	89.5	0.819	85.7	81.1
FIBA	0.750	85.7	72.2	0.842	85.7	84.2	0.797	85.7	78.4
K2C6A	0.746	85.7	72.2	0.883	100.0	79.0	0.817	100.0	74.5
TFR1	0.734	85.7	72.2	0.880	85.7	89.5	0.809	85.7	81.1
K1C9	0.722	71.4	83.3	0.714	78.6	63.2	0.718	71.4	73.0
HV305	0.718	64.3	88.9	0.797	57.1	100.0	0.759	57.1	94.6
RELN	0.718	92.9	55.6	0.722	64.3	84.2	0.72	64.3	81.1
KAIN	0.853	78.6	88.9	0.919	71.4	94.7	0.887	78.6	86.5
KLKB1	0.829	78.6	83.3	0.823	78.6	73.7	0.826	78.6	78.4
APOL1	0.825	85.7	66.7	0.786	64.3	94.7	0.805	64.3	89.2
HEMO	0.806	78.6	72.2	0.906	78.6	94.7	0.857	78.6	83.8
APOA2	0.798	92.9	66.7	0.786	78.6	68.4	0.792	92.9	56.8
APOA1	0.790	45.8	72.2	0.782	64.3	100.0	0.786	64.3	91.9
AMBP	0.782	50.0	100.0	0.774	85.7	57.9	0.778	85.7	59.5
APOF	0.758	78.6	83.3	0.876	78.6	89.5	0.819	78.6	86.5
VTDB	0.750	64.3	88.9	0.786	57.1	94.7	0.768	57.1	94.6
ALBU	0.738	42.9	100.0	0.936	92.9	89.5	0.84	92.9	62.2
ITIH1	0.738	85.7	61.1	0.831	85.7	73.7	0.786	85.7	67.6
ITIH2	0.730	85.7	66.7	0.789	85.7	68.4	0.761	85.7	67.6
RAB10	0.710	92.9	50.0	0.726	92.9	57.9	0.718	92.9	54.1

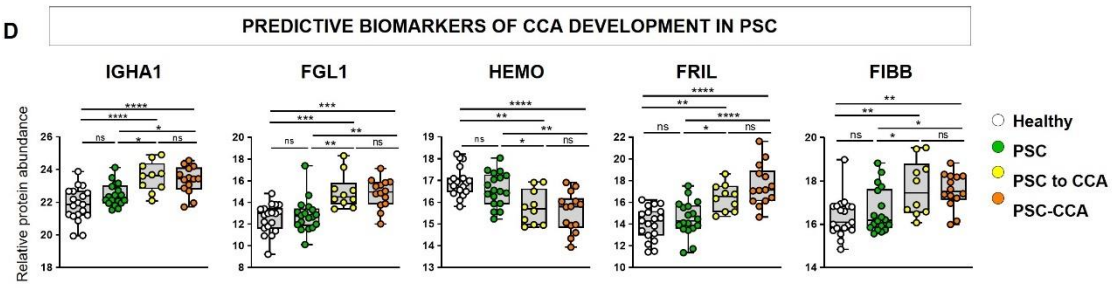
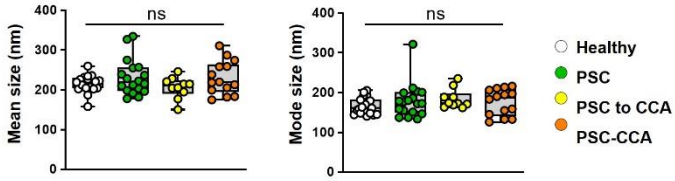


Figure 4.2. Isolated serum EV fractions are enriched in exosomes and microvesicles, and contain proteins with high diagnostic capacity for PSC-CCA. (**A**) Characterization of serum EVs by TEM, NTA and immunoblot. (**B**) Volcano plot of identified proteins in PSC-CCA *vs* healthy individuals, PSC-CCA *vs* PSC and in PSC-CCA *vs* a control group of healthy individuals and patients with PSC (non-malignant control group). Significantly-enriched proteins are colored in red and proteins with lower abundance in blue. (**C**) Heatmap and table with diagnostic values of EV-proteins altered in patients with concomitant PSC-CCA compared to the non-malignant control group. (**D**) Box plots of biomarkers of risk for CCA development in PSC.

A Nanoparticle Tracking Analysis (NTA)



B

Protein	PSC-CCA (n=14) *vs* PSC (n=18)			PSC-CCA (n=14) *vs* Healthy (n=19)			PSC-CCA (n=14) *vs* [Healthy + PSC] (n=37)		
	AUC (95% CI)	SEN, % (95% CI)	SPE, % (95% CI)	AUC (95% CI)	SEN, % (95% CI)	SPE, % (95% CI)	AUC (95% CI)	SEN, % (95% CI)	SPE, % (95% CI)
FRIL	**0.909** (0.81, 1.00)	**85.7** (57, 98)	**88.9** (65, 99)	**0.951** (0.88, 1.00)	**85.7** (57, 98)	**94.7** (74, 100)	**0.931** (0.86, 1.00)	**85.7** (57, 98)	**91.9** (78, 98)
LAC2	**0.905** (0.80,1.00)	**85.7** (57, 98)	**83.3** (59, 96)	**0.759** (0.59, 0.93)	**57.1** (29, 82)	**89.5** (67, 99)	**0.830** (0.71, 0.95)	**57.1** (29, 82)	**94.6** (82, 99)
FGL1	**0.869** (0.73, 1.00)	**85.7** (57, 98)	**88.9** (65, 99)	**0.895** (0.78, 1.00)	**78.6** (49, 95)	**94.7** (74, 100)	**0.882** (0.77, 0.99)	**78.6** (49, 95)	**91.9** (78, 98)
LV301	**0.853** (0.72, 0.99)	**85.7** (57, 98)	**83.3** (59, 96)	**0.904** (0.80, 1.00)	**92.9** (66, 100)	**79.0** (54, 94)	**0.879** (0.78, 0.98)	**85.7** (57, 98)	**83.8** (68, 94)
OIT3	**0.849** (0.72, 0.98)	**64.3** (35, 87)	**94.4** (73, 100)	**0.951** (0.89, 1.00)	**92.9** (66, 100)	**84.2** (60, 97)	**0.902** (0.81, 0.99)	**78.6** (49, 95)	**83.8** (68, 94)
HV308	**0.841** (0.70, 0.98)	**85.7** (57, 98)	**77.8** (52, 94)	**0.906** (0.81, 1.00)	**92.9** (66, 100)	**73.7** (49, 91)	**0.875** (0.78, 0.97)	**85.7** (57, 98)	**75.7** (59, 88)
SVEP1	**0.825** (0.68, 0.97)	**100** (77, 100)	**50.0** (26, 74)	**0.763** (0.60, 0.93)	**78.6** (49, 95)	**73.7** (49, 91)	**0.793** (0.67, 0.92)	**92.9** (66, 100)	**56.8** (40, 73)
LFG3	**0.796** (0.63, 0.96)	**78.6** (79, 83)	**83.3** (59, 96)	**0.782** (0.61, 0.96)	**78.6** (49, 95)	**84.2** (60, 97)	**0.789** (0.64, 0.94)	**78.6** (49, 95)	**83.8** (68, 94)
IGHA1	**0.790** (0.62, 0.96)	**64.3** (35, 87)	**88.9** (65, 99	**0.876** (0.75, 1.00)	**85.7** (57, 98)	**84.2** (60, 97)	**0.834** (0.70, 0.96)	**85.7** (57, 98)	**75.7** (59, 89)
K22E	**0.790** (0.63, 0.95)	**92.9** (66, 100)	**66.7** (41, 87)	**0.722** (0.54, 0.90)	**92.9** (66, 100)	**63.2** (38, 84)	**0.755** (0.62, 0.89)	**92.9** (66, 100)	**64.9** (48, 80)
LV403	**0.786** (0.62, 0.95)	**71.4** (42, 92)	**77.8** (52, 94)	**0.733** (0.56, 0.91)	**92.9** (66, 100)	**57.9** (34, 80)	**0.759** (0.63, 0.89)	**92.9** (66, 100)	**56.8** (40, 73)
PIGR	**0.782** (0.61, 0.96)	**85.7** (57, 98)	**77.8** (52, 94)	**0.992** (0.97, 1.00)	**100** (77, 100)	**94.7** (74, 100)	**0.890** (0.80, 0.98)	**85.7** (57, 98)	**89.2** (75, 97)
FCGBP	**0.782** (0.62,0.94)	**85.7** (57, 98)	**72.2** (47, 90)	**0.951** (0.88, 1.00)	**85.7** (57, 98)	**94.7** (74, 100)	**0.869** (0.77, 0.97)	**85.7** (57, 99)	**83.8** (68, 94)
FIBG	**0.778** (0.60, 0.96)	**85.7** (57, 98)	**77.8** (52, 94)	**0.797** (0.62, 0.97)	**85.7** (57, 98)	**79.1** (54, 94)	**0.788** (0.62, 0.95)	**85.7** (57, 98)	**78.4** (62, 90)
IGLL5	**0.770** (0.60, 0.94)	**78.6** (49, 95)	**72.2** (47, 90)	**0.823** (0.68, 0.97)	**64.3** (35, 87)	**89.5** (67, 99)	**0.797** (0.66, 0.93)	**64.3** (35, 87)	**83.8** (68, 94)
FIBB	**0.758** (0.59, 0.93)	**85.7** (57, 98)	**72.2** (47, 90)	**0.876** (0.74, 1.00)	**85.7** (57, 98)	**89.5** (67, 99)	**0.819** (0.70, 0.94)	**85.7** (57, 98)	**81.1** (65, 92)
FIBA	**0.750** (0.57, 0.93)	**85.7** (57, 98)	**72.2** (47, 90)	**0.842** (0.70, 0.99)	**85.7** (57, 98)	**84.2** (60, 97)	**0.797** (0.66, 0.94)	**85.7** (57, 98)	**78.4** (62, 90)
K2C6A	**0.746** (0.57, 0.92)	**85.7** (57, 98)	**72.2** (47, 90)	**0.883** (0.77, 1.00)	**100** (77, 100)	**79.0** (54, 94)	**0.817** (0.70, 0.93)	**100** (77, 100)	**64.86** (48, 80)
TFR1	**0.734** (0.55, 0.92)	**85.7** (57, 98)	**72.2** (47, 90)	**0.880** (0.74, 1.00)	**85.7** (57, 98)	**89.5** (67, 99)	**0.809** (0.66, 0.96)	**85.7** (57, 98)	**81.1** (65, 92)
K1C9	**0.722** (0.53, 0.91)	**71.4** (42, 92)	**83.3** (59, 96)	**0.714** (0.53, 0.90)	**78.6** (49, 95)	**63.2** (38, 84)	**0.718** (0.55, 0.88)	**71.4** (42, 92)	**73.0** (56, 86)
HV305	**0.718** (0.53, 0.91)	**64.3** (34, 87)	**88.9** (65, 99)	**0.797** (0.64, 0.96)	**57.1** (29, 82)	**100** (82, 100)	**0.759** (0.60, 0.92)	**57.1** (29, 82)	**94.6** (82, 99)
RELN	**0.718** (0.53, 0.90)	**92.9** (66, 100)	**55.6** (31, 79)	**0.722** (0.53, 0.91)	**64.3** (35, 87)	**84.2** (60, 97)	**0.720** (0.56, 0.89)	**64.3** (35, 87)	**81.1** (65, 92)
KAIN	**0.853** (0.71, 0.99)	**78.6** (49, 95)	**88.9** (65, 99)	**0.919** (0.83, 1.00)	**71.4** (42, 92)	**94.7** (74, 100)	**0.887** (0.80, 0.98)	**78.6** (49, 95)	**86.5** (71, 96)
KLKB1	**0.829** (0.69, 0.97)	**78.6** (49, 95)	**83.3** (59, 96)	**0.823** (0.68, 0.97)	**78.6** (49, 95)	**73.7** (49, 91)	**0.826** (0.70, 0.95)	**78.6** (49, 95)	**78.4** (62, 90)
APOL1	**0.825** (0.68, 0.97)	**85.7** (57, 98)	**66.7** (41, 87)	**0.786** (0.61, 0.96)	**64.3** (35, 87)	**94.7** (74, 100)	**0.805** (0.66, 0.95)	**64.3** (35, 87)	**89.2** (75, 97)
HEMO	**0.806** (0.66, 0.96)	**78.6** (49, 95)	**72.2** (47, 90)	**0.906** (0.80, 1.00)	**78.6** (49, 95)	**94.7** (74, 100)	**0.857** (0.74, 0.97)	**78.6** (49, 95)	**83.8** (68, 94)
APOA2	**0.798** (0.64, 0.95)	**92.9** (66, 100)	**66.7** (41, 87)	**0.786** (0.63, 0.94)	**78.6** (49, 95)	**68.4** (44, 87)	**0.792** (0.66, 0.93)	**92.9** (66, 100)	**56.8** (40, 73)
APOA1	**0.790** (0.63, 0.95)	**78.6** (49, 95)	**72.2** (47, 90)	**0.782** (0.60, 0.96)	**64.3** (35, 87)	**100** (82, 100)	**0.786** (0.62, 0.95)	**64.3** (35, 87)	**91.9** (78, 98)
AMBP	**0.782** (0.61, 0.95)	**50.0** (23, 77)	**100** (82, 100)	**0.774** (0.61, 0.94)	**85.7** (57, 98)	**57.9** (34, 80)	**0.778** (0.62, 0.93)	**85.7** (57, 98)	**59.5** (42, 75)
APOF	**0.758** (0.58, 0.94)	**78.6** (49, 95)	**83.3** (59, 96)	**0.876** (0.75, 1.00)	**78.6** (49, 95)	**89.5** (67, 99)	**0.819** (0.68, 0.95)	**78.6** (49, 95)	**86.5** (71, 96)
VTDB	**0.750** (0.57, 0.93)	**64.3** (35, 87)	**88.9** (65, 99)	**0.786** (0.62, 0.95)	**57.1** (29, 82)	**94.7** (74, 100)	**0.768** (0.61, 0.93)	**57.1** (29, 82)	**94.6** (82, 99)
ALBU	**0.738** (0.56, 0.92)	**42.9** (18, 71)	**100** (82, 100)	**0.936** (0.85, 1.00)	**92.9** (66, 100)	**89.5** (67, 99)	**0.840** (0.72, 0.96)	**92.9** (66, 100)	**62.2** (45, 78)
ITIH1	**0.738** (0.56, 0.92)	**85.7** (57, 98)	**61.1** (36, 83)	**0.831** (0.67, 0.99)	**85.7** (57, 98)	**73.7** (49, 91)	**0.786** (0.63, 0.94)	**85.7** (57, 98)	**67.6** (50, 82)
ITIH2	**0.730** (0.55, 0.91)	**85.7** (57, 98)	**66.7** (41, 87)	**0.789** (0.63, 0.95)	**85.7** (57, 98)	**68.4** (44, 87)	**0.761** (0.62, 0.91)	**85.7** (57, 98)	**67.6** (50, 82)
RAB10	**0.710** (0.52, 0.90)	**92.9** (66, 100)	**50.0** (26, 74)	**0.726** (0.54, 0.91)	**92.9** (66, 100)	**57.9** (34, 80)	**0.718** (0.56, 0.88)	**92.9** (66, 100)	**54.1** (37, 71)

C PREDICTIVE BIOMARKERS OF CCA DEVELOPMENT IN PSC

Protein	PSC to CCA (n=10) *vs* PSC (n=18)					
	AUC (95% CI)	SEN, % (95% CI)	SPE, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	AI, % (95% CI)
FGL1	**0.900** (0.78, 1.00)	**100.0** (69, 100)	**77.8** (52, 94)	**71.4** (51, 86)	**100.0**	**85.7** (67, 96)
IGHA1	**0.833** (0.68, 0.99)	**90.0** (56, 100)	**66.7** (41, 87)	**60.0** (43, 75)	**92.3** (64, 99)	**75.0** (55, 89)
FRIL	**0.828** (0.68, 0.98)	**90.0** (56, 100)	**66.7** (41, 87)	**60.0** (43, 75)	**92.3** (64, 99)	**75.0** (55, 89)
FIBB	**0.789** (0.62, 0.96)	**90.0** (56, 100)	**66.7** (41, 87)	**60.0** (43, 75)	**92.3** (64, 99)	**75.0** (55, 89)
LAC2	**0.767** (0.57, 0.97)	**70.0** (35, 93)	**88.9** (65, 99)	**77.8** (47, 93)	**84.2** (67, 93)	**82.1** (63, 94)
FIBG	**0.756** (0.56, 0.95)	**90.0** (56, 100)	**61.1** (36, 83)	**56.3** (41, 70)	**91.7** (62, 99)	**71.4** (51, 87)
FIBA	**0.733** (0.54, 0.92)	**70.0** (35, 93)	**72.2** (47, 90)	**58.3** (37, 77)	**81.3** (62, 92)	**71.4** (51, 87)
PIGR	**0.733** (0.54, 0.93)	**80.0** (44, 98)	**77.8** (52, 94)	**66.7** (44, 83)	**87.5** (66, 96)	**78.6** (59, 92)
HEMO	**0.778** (0.59, 0.96)	**70.0** (35, 93)	**83.3** (59, 96)	**70.0** (43, 88)	**83.3** (65, 93)	**78.6** (59, 92)

Figure 4.3. Proteomic characterization of serum extracellular vesicles: related to Figure 4.2.

3.2 Serum EVs contain protein biomarkers for the early, accurate and etiology-based diagnosis of CCAs, with higher accuracy than CA19-9

In order to evaluate whether the aforementioned candidate biomarkers for CCAs are specific for patients with PSC background or have a common diagnostic ability also for CCAs of other etiologies, these results were combined with an analogous EV-proteomic dataset that includes patients with CCAs from non-PSC etiologies (PRIDE accession: PXD026199) (**Figure 4.4**).

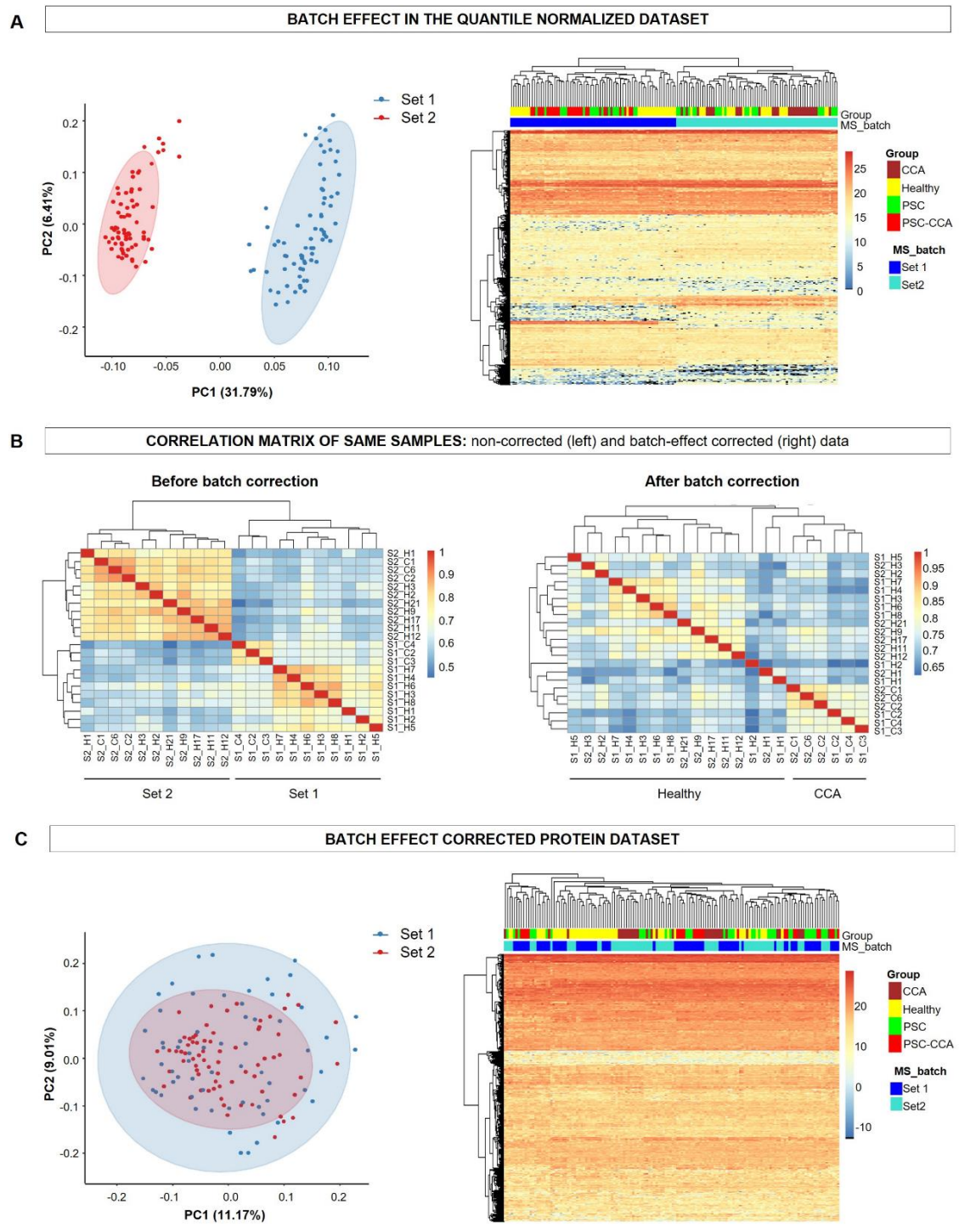
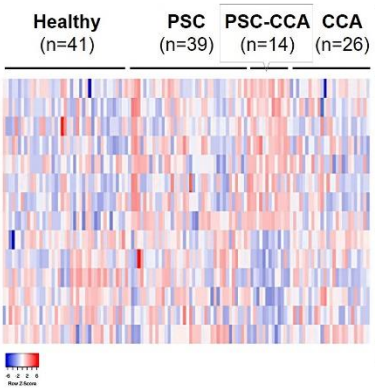


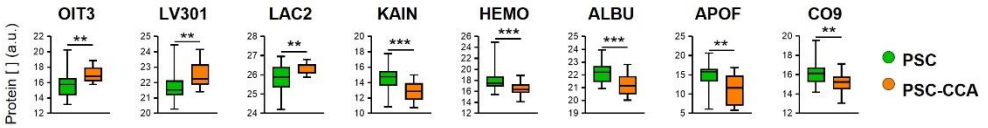
Figure 4.4. Batch effect correction of two serum EV mass spectrometry-based proteomic sets. (**A**) Diagnostics of batch effects in quantile normalized peptide data by principal component analysis (PCA) and by a heatmap visualizing the hierarchical clustering dendrogram. (**B**) Quality control on batch-corrected data matrix illustrated by correlation plot heatmaps of the replicates before and after batch correction. (**C**) Batch effect correction confirmation after protein quantification by PCA and by hierarchical clustering dendrogram illustrating heatmap.

The analysis of the batch-corrected EV proteome resulted in the identification of a total of 242 proteins, among which the abundance of 40 differs according to the presence/absence of PSC. Taking them into account, 14 proteins allowed the differential diagnosis of patients with PSC-CCA vs isolated PSC with greater accuracy than serum CA19-9 (AUC=0.675), harboring KAIN, HEMO and OIT3 the highest diagnostic capacity (AUC=0.846, 0,815 and 0.788, respectively) (**Figure 4.5A; Figure 4.6A**). Furthermore, among other proteins, decreased HEMO levels and increased LV301 levels were linked to the greatest risk of CCA occurrence in patients with PSC [OR 33.3 (95% CI 5.5 – 200.3) and OR 20.8 (95% CI 2.5 – 175.7), respectively], while increased serum levels of CA19-9 were not indicative of greater risk of CCA presence [OR 3.5 (95% CI 0.9 – 13.2). On the other hand, the abundance of 6 proteins allowed the specific diagnosis of non-PSC CCAs, providing AUC values up to 0.874 when compared with healthy controls (**Figure 4.5B; Figure 4.6B**). Considering EV-proteins whose levels were not associated with the presence/absence of PSC, the abundance of 23 enabled the diagnosis of patients with CCA regardless etiology (pan-CCA) (**Figure 4.5C; Figure 4.6C**). Among them, PIGR, FIBG and APOA1 stood out as the best individual diagnostic biomarkers, displaying AUC values of 0.812, 0.803 and 0.787, respectively, when comparing all CCA samples with the “non-malignant control group”. Noteworthy, increased FIBG levels and decreased APOA1 levels provided the highest risk of CCA occurrence [OR 14.5 (95% CI 5.7 – 37.1) and OR 13.1 (95% CI 5.1 – 33.8), respectively].

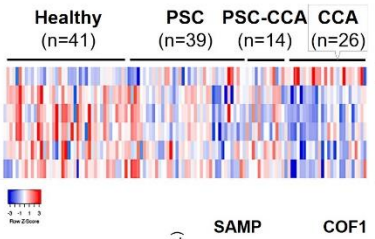
A **Specific PSC-CCA biomarkers**



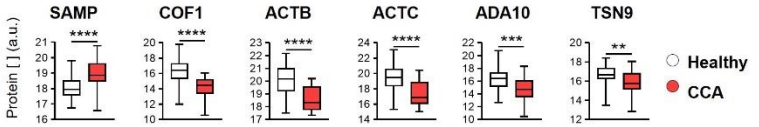
Protein	PSC-CCA (n=14) *vs* PSC (n=39)						
	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)	OR (95% CI)
OIT3	0.788	85.7	64.1	46.2	92.6	69.8	**10.7** (2.1 – 54.9)
LV301	0.766	92.9	61.5	46.4	96.0	69.8	**20.8** (2.5 – 175.7)
K22E	0.755	78.6	69.2	47.8	90.0	71.7	**8.3** (1.9 – 35.0)
HV308	0.742	78.6	64.1	44.0	89.3	67.9	**6.5** (1.6 – 27.5)
LAC2	0.742	85.7	61.5	44.4	92.3	67.9	**9.6** (1.9 – 49.0)
TFR1	0.716	85.7	69.2	50.0	93.1	73.6	**13.5** (2.6 – 69.9)
IGLL5	0.690	64.3	74.4	47.4	85.3	71.7	**5.2** (1.4 – 19.3)
PLCH1	0.690	92.9	48.7	39.4	95.0	60.4	**12.4** (1.5 – 103.7)
KAIN	0.846	71.4	87.2	66.7	89.5	83.0	**17.0** (3.8 – 75.6)
HEMO	0.815	64.3	94.9	81.8	88.1	86.8	**33.3** (5.5 – 200.3)
ALBU	0.788	57.1	92.3	72.7	85.7	83.0	**16.0** (3.3 – 78.0)
APOF	0.749	78.6	71.8	50.0	90.3	73.6	**9.3** (2.2 – 40.0)
CO9	0.742	71.4	74.4	50.0	87.9	73.6	**7.3** (1.9 – 28.4)
A2MG	0.718	64.3	84.6	60.0	86.8	79.3	**9.9** (2.4 – 40.0)
CA19-9	0.675	46.2	97.2	85.7	83.3	83.7	**3.5** (0.9 – 13.2)



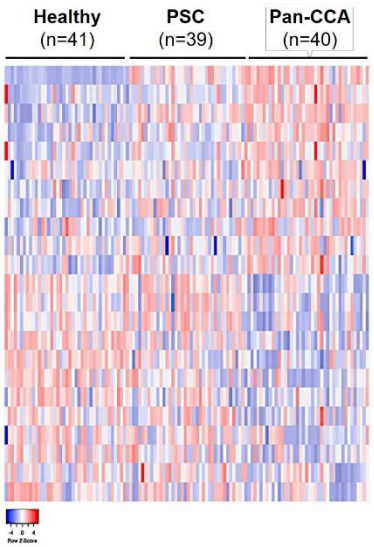
B **Non-PSC CCA biomarkers**



Protein	CCA (n=26) *vs* Healthy (n=41)						
	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)	OR (95% CI)
SAMP	0.832	88.5	70.7	65.7	90.6	77.6	**18.5** (4.7 – 73.5)
COF1	0.874	80.8	80.5	72.4	86.8	80.6	**17.3** (5.0 – 60.1)
ACTB	0.848	65.4	90.2	81.0	80.4	80.6	**17.5** (4.7 – 64.8)
ACTC	0.830	65.4	90.2	81.0	80.4	80.6	**17.5** (4.7 - 64.8)
ADA10	0.746	53.9	87.8	73.7	75.0	74.6	**8.4** (2.5 – 28.2)
TSN9	0.720	65.4	82.9	70.8	79.1	76.1	**9.2** (2.9 – 28.9)



C **Pan-CCA biomarkers**



Protein	Pan-CCA (n=40) ***vs*** [Healthy + PSC] (n=80)						
	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)	OR (95% CI)
PIGR	0.812	77.5	77.7	63.3	87.3	77.5	**11.9** (4.8 – 29.4)
FIBG	0.803	67.5	87.5	73.0	84.3	80.8	**14.5** (5.7 – 37.1)
PPBT	0.760	70.0	72.5	56.0	82.9	71.7	**6.2** (2.7 – 14.2)
A1AT	0.759	67.5	76.3	58.7	82.4	73.3	**6.7** (2.9 – 15.4)
FIBB	0.717	67.5	71.3	54.0	81.4	70.0	**5.1** (2.3 – 11.7)
RELN	0.712	60.0	80.0	60.0	80.0	73.3	**6.0** (2.6 – 13.9)
VWF	0.700	70.0	72.5	56.0	82.9	71.7	**6.2** (2.7 – 14.2)
IGHG4	0.690	65.0	66.3	49.1	79.1	65.8	**3.6** (1.6 – 8.1)
LRP1	0.686	60.0	77.5	57.1	79.5	71.7	**5.2** (2.3 – 11.8)
FGL1	0.686	42.5	90.0	68.0	75.8	74.2	**6.7** (2.5 – 17.4)
IGHA1	0.682	55.0	75.0	52.4	76.9	68.3	**3.7** (1.6 – 8.2)
APOA1	0.787	62.5	88.8	73.5	82.6	80.0	**13.1** (5.1 – 33.8)
HPTR	0.749	52.5	87.5	67.7	78.7	75.8	**7.7** (3.1 – 19.2)
APOL1	0.747	62.5	83.8	65.8	81.7	76.7	**8.6** (3.6 – 20.6)
KLKB1	0.742	82.5	63.8	53.3	87.9	70.0	**8.3** (3.3 – 21.1)
GELS	0.738	82.5	63.8	53.2	87.9	70.0	**8.3** (3.3 – 21.1)
FLNA	0.732	52.5	86.3	65.6	78.4	75.0	**6.9** (2.9 – 16.9)
ACTN1	0.705	77.5	62.5	50.8	84.8	67.5	**5.7** (2.4 – 13.7)
ITIH2	0.687	70.0	66.3	50.9	81.5	67.5	**4.6** (2.0 – 10.4)
1433Z	0.687	70.0	71.3	54.9	82.6	70.8	**5.8** (2.5 – 13.3)
TLN1	0.686	92.5	40.0	43.5	91.4	57.5	**8.2** (2.3 – 29.0)
VTDB	0.662	37.5	93.8	75.0	75.0	75.0	**9.0** (3.0 – 27.3)
IMA8	0.659	92.5	35.0	41.6	90.3	54.2	**6.6** (1.9 – 23.5)

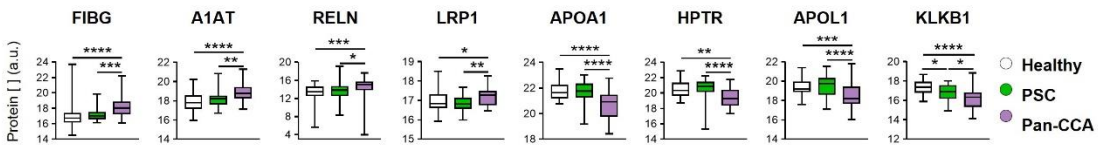


Figure 4.5. Serum EV-protein biomarkers for the diagnosis of CCA according to tumor etiology. Biomarkers for the specific diagnosis of (**A**) CCA in patients with PSC (specific PSC-CCA biomarkers), **(B**) CCA in patients without PSC (non-PSC CCA biomarkers) and (**C**) CCA regardless etiology (Pan-CCA biomarkers). Significantly-enriched proteins are colored in red and proteins with lower abundance in blue.

A **Specific PSC-CCA biomarkers**

Protein	**PSC-CCA** (n=14) *vs* **Healthy** (n=41)						**PSC-CCA** (n=14) *vs* **CCA** (n=26)						**PSC-CCA** (n=14) *vs* **[Healthy + PSC + CCA]** (n=106)					
	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)
OIT3	0.895	100.0	70.7	53.9	100.0	78.2	0.904	92.9	76.9	68.4	95.2	82.5	0.858	92.9	68.9	28.3	98.7	71.7
LV301	0.897	78.6	87.8	68.8	92.3	85.5	0.775	92.9	65.4	59.1	94.4	75.0	0.819	92.9	67.0	27.1	98.6	70.0
K22E	0.702	92.9	53.7	40.6	95.7	63.6	0.838	85.7	76.9	66.7	90.9	80.0	0.755	92.9	56.6	22.0	98.4	60.8
HV308	0.892	78.6	90.2	73.3	92.5	87.3	0.838	78.6	80.8	68.8	87.5	80.0	0.823	78.6	78.3	32.4	96.5	78.3
LAC2	0.801	57.1	92.7	72.7	86.4	83.6	0.698	100.0	50.0	51.9	100.0	67.5	0.754	85.7	60.4	22.2	97.0	63.3
TFR1	0.834	85.7	82.9	63.2	94.4	83.6	0.783	85.7	76.9	66.7	90.9	80.0	0.778	85.7	76.4	32.4	97.6	77.5
IGLL5	0.817	64.3	87.8	64.3	87.8	81.8	0.816	71.4	84.6	71.4	84.6	80.0	0.770	71.4	76.4	28.6	95.3	75.8
PLCH1	0.889	92.9	82.9	65.0	97.1	85.5	0.799	92.9	76.9	68.4	95.2	82.5	0.793	92.9	68.9	28.3	98.7	71.7
KAIN	0.868	85.7	80.5	60.0	94.3	81.8	0.934	85.7	92.3	85.7	92.3	90.0	0.876	85.7	80.2	36.4	97.7	80.8
HEMO	0.756	64.3	85.4	60.0	87.5	80.0	0.808	78.6	88.5	78.6	88.5	85.0	0.790	71.4	84.0	37.0	95.7	82.5
ALBU	0.848	92.9	68.3	50.0	96.6	74.6	0.725	57.1	88.5	72.7	79.3	77.5	0.796	57.1	92.5	50.0	94.2	88.3
APOF	0.805	57.1	97.6	88.9	87.0	87.3	0.714	50.0	96.2	87.5	78.1	80.0	0.762	78.6	68.9	25.0	96.1	70.0
CO9	0.829	71.4	87.8	66.7	90.0	83.6	0.745	71.4	80.8	66.7	84.0	77.5	0.776	71.4	81.1	33.3	95.6	80.0
A2MG	0.693	64.3	80.5	52.9	86.8	76.4	0.777	64.3	96.2	90.0	83.3	85.0	0.723	64.3	85.9	37.5	94.8	83.3

B **Non-PSC CCA biomarkers**

Protein	**CCA** (n=26) *vs* **PSC** (n=39)						**CCA** (n=26) *vs* **PSC-CCA** (n=14)						**CCA** (n=26) *vs* **[Healthy + PSC + PSC-CCA]** (n=94)					
	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)
SAMP	0.740	69.2	74.4	64.3	78.4	72.3	0.720	80.8	64.3	80.8	64.3	75.0	0.777	88.5	59.6	37.7	94.9	65.8
COF1	0.702	84.6	56.4	56.4	84.6	67.7	0.709	100.0	50.0	78.8	100.0	82.5	0.778	100.0	50.0	35.6	100.0	60.8
ACTB	0.723	53.9	89.7	77.8	74.5	75.4	0.725	46.2	92.9	92.3	48.2	62.5	0.778	53.9	90.4	60.9	87.6	82.5
ACTC	0.646	53.9	79.5	63.6	72.1	69.2	0.695	50.0	85.7	86.7	48.0	62.5	0.734	53.9	86.2	51.9	87.1	79.2
ADA10	0.662	53.9	79.5	63.6	72.1	69.2	0.772	46.2	100.0	100.0	50.0	65.0	0.715	53.9	84.0	48.3	86.8	77.5
TSN9	0.702	65.4	74.4	63.0	76.3	70.8	0.706	65.4	85.7	89.5	57.1	72.5	0.710	65.4	79.8	47.2	89.3	76.7

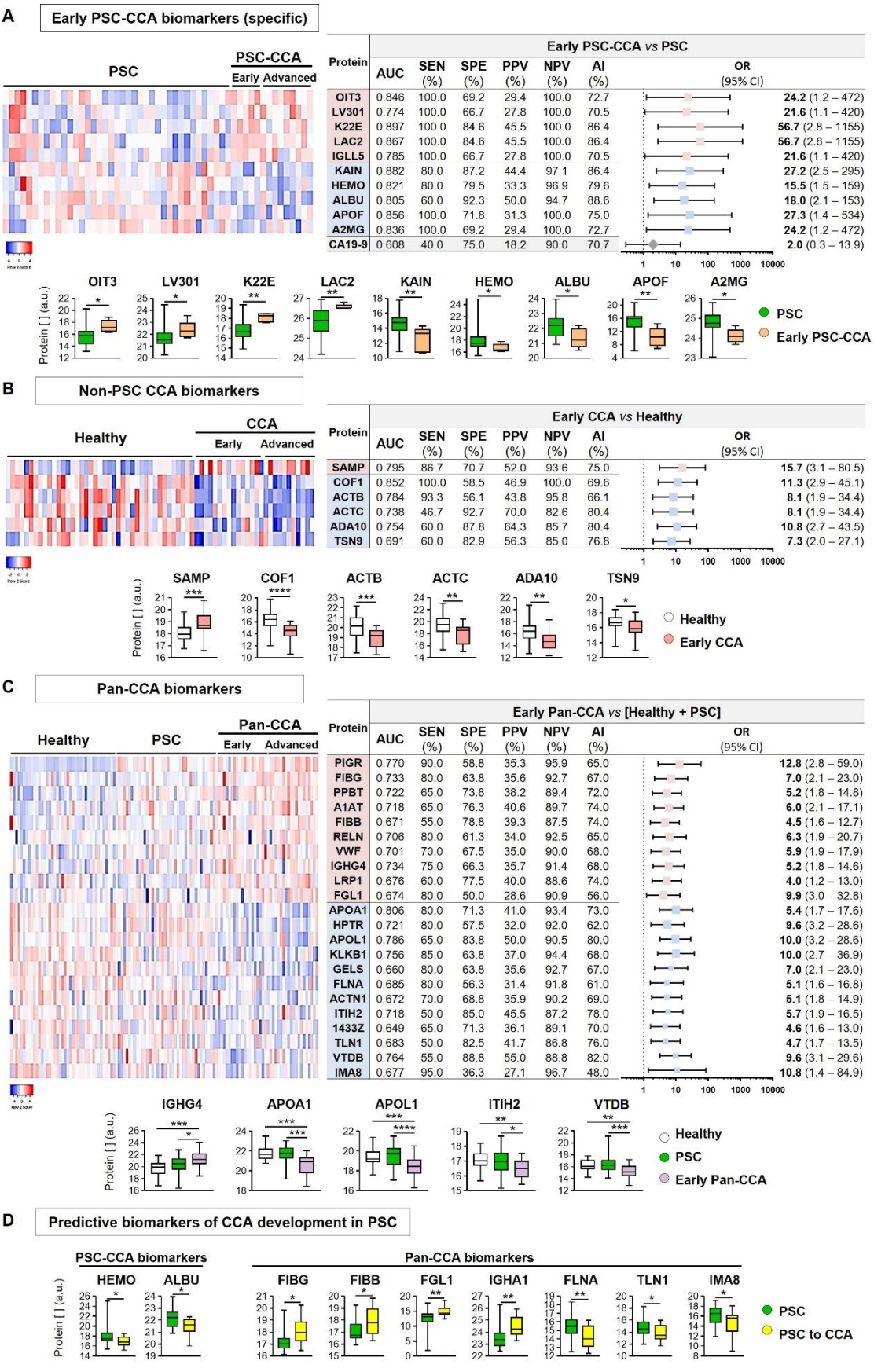
C **Pan-CCA biomarkers**

Protein	**Pan-CCA** (n=40) *vs* **Healthy** (n=41)						**Pan-CCA** (n=40) *vs* **PSC** (n=39)					
	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)
PIGR	0.974	97.5	92.7	92.9	97.4	95.1	0.642	75.0	59.0	65.2	69.7	67.1
FIBG	0.826	90.0	65.9	72.0	87.1	77.8	0.778	67.5	87.2	84.4	72.3	77.2
PPBT	0.872	87.5	78.1	79.6	86.5	82.7	0.643	55.0	74.4	68.8	61.7	64.6
A1AT	0.793	95.0	58.5	69.1	92.3	76.5	0.722	67.5	76.9	75.0	69.8	72.2
FIBB	0.766	67.5	78.1	75.0	71.1	72.8	0.665	77.5	56.4	64.6	71.0	67.1
RELN	0.738	60.0	82.9	77.4	68.0	71.6	0.686	60.0	76.9	72.7	65.2	68.4
VWF	0.754	72.5	78.1	76.3	74.4	75.3	0.643	60.0	74.4	70.6	64.4	67.1
IGHG4	0.740	65.0	75.6	72.2	68.9	70.4	0.638	40.0	84.6	72.7	57.9	62.0
LRP1	0.665	65.0	68.3	66.7	66.7	66.7	0.708	60.0	82.1	77.4	66.7	70.9
FGL1	0.691	42.5	92.7	85.0	62.3	67.9	0.680	55.0	79.5	73.3	63.3	67.1
IGHA1	0.733	57.5	80.5	74.2	66.0	69.1	0.628	47.5	76.9	67.9	58.8	62.0
APOA1	0.797	62.5	92.7	89.3	71.7	77.8	0.776	65.0	84.6	81.3	70.2	74.7
HPTR	0.718	57.5	80.5	74.2	66.0	69.1	0.781	82.5	69.2	73.3	79.4	76.0
APOL1	0.741	62.5	90.2	86.2	71.2	76.5	0.753	80.0	66.7	71.1	76.5	73.4
KLKB1	0.815	82.5	73.2	75.0	81.1	77.8	0.665	82.5	53.9	64.7	75.0	68.4
GELS	0.829	85.0	80.5	81.0	84.6	82.7	0.643	62.5	66.7	65.8	63.4	64.6
FLNA	0.798	57.5	92.7	88.5	69.1	75.3	0.662	45.0	87.2	78.3	60.7	65.8
ACTN1	0.756	82.5	70.7	73.3	80.6	76.5	0.651	70.0	61.5	65.1	66.7	65.8
ITIH2	0.732	70.0	75.6	73.7	72.1	72.8	0.640	67.5	59.0	62.8	63.9	63.3
1433Z	0.688	70..0	73.2	71.8	71.4	71.6	0.685	70.0	69.2	70.0	69.2	69.6
TLN1	0.738	92.5	48.8	63.8	87.0	70.4	0.632	35.0	89.7	77.8	57.4	62.0
VTDB	0.658	37.5	95.1	88.2	60.9	66.7	0.665	35.0	94.9	87.5	58.7	64.6
IMA8	0.680	92.5	39.0	59.7	84.2	65.4	0.637	62.5	64.1	64.1	62.5	63.3

Figure 4.6: related to Figure 4.5. Diagnostic test evaluation of specific PSC-CCA biomarkers (**A**), non-PSC CCA biomarkers (**B**) and Pan-CCA biomarkers (**C**).

Most of the previously identified single candidate biomarkers retained their diagnostic accuracy when considering only patients with early-stage CCA. For instance, while CA19-9 levels failed to differentiate patients with early-stage PSC-CCA from patients with isolated PSC, higher levels of K22E and LAC2 increased the odds for CCA occurrence by 56.7-fold, reinforcing the utility of these biomarkers for the early diagnosis of CCA (**Figure 4.7A-C; Figure 4.8A-C**). Interestingly, some PSC-CCA or pan-CCA biomarkers were already found altered in patients with PSC who progressed to CCA during follow-up even when no clinical signs of tumor development were evident (**Figure 4.7D; Figure 4.8D**). Among PSC-CCA biomarkers, the levels of ALBU (AUC=0.726) and HEMO (AUC=0.705) were already found altered in the PSC to CCA group compared to the isolated PSC group, an event that was not found for CA19.9 (AUC=0.500), thus pinpointing their potential clinical utility to predict bile duct cancer development in patients with PSC [HEMO (PPV 42.9% and NPV 88.6%) and ALBU (PPV 33.3% and NPV 100.0%)]. Similarly, among pan-CCA biomarkers, the levels of FIBG, FIBB, FGL1, IGHA1, FLNA, TLN1 and IMA8 might also be considered as predictors of CCA development in individuals with PSC, with AUC values up to 0.787 and PPV and NPVs of up to 54.6 and 96.4, respectively), identifying FLNA as the most accurate predictive biomarker (**Figure 4.7D; Figure 4.8D**).

Figure 4.7. Serum EV-protein biomarkers for the early diagnosis of CCA according to tumor etiology. Early biomarkers for the specific diagnosis of (**A**) CCA in patients with PSC (specific PSC-CCA biomarkers), (**B**) CCA in patients without PSC (non-PSC CCA biomarkers) and (**C**) CCA regardless etiology (Pan-CCA biomarkers). Significantly-enriched proteins are colored in red and proteins with lower abundance in blue. (**D**) Box plots of biomarkers that predict CCA development in patients with PSC.


A
Early PSC-CCA biomarkers (specific)
PSC
PSC-CCA
Early
Advanced
Early PSC-CCA vs PSC
Protein
AUC
SEN (%)
SPE (%)
PPV (%)
NPV (%)
AI (%)
OR (95% CI)
OIT3 0.846 100.0 69.2 29.4 100.0 72.7 24.2 (1.2 – 472)
LV301 0.774 100.0 66.7 27.8 100.0 70.5 21.6 (1.1 – 420)
K22E 0.897 100.0 84.6 45.5 100.0 86.4 56.7 (2.8 – 1155)
LAC2 0.867 100.0 84.6 45.5 100.0 86.4 56.7 (2.8 – 1155)
IGLL5 0.785 100.0 66.7 27.8 100.0 70.5 21.6 (1.1 – 420)
KAIN 0.882 80.0 87.2 44.4 97.1 86.4 27.2 (2.5 – 295)
HEMO 0.821 80.0 79.5 33.3 96.9 79.6 15.5 (1.5 – 159)
ALBU 0.805 60.0 92.3 50.0 94.7 88.6 18.0 (2.1 – 153)
APOF 0.856 100.0 71.8 31.3 100.0 75.0 27.3 (1.4 – 534)
A2MG 0.836 100.0 69.2 29.4 100.0 72.7 24.2 (1.2 – 472)
CA19-9 0.608 40.0 75.0 18.2 90.0 70.7 2.0 (0.3 – 13.9)
1 10 100 1000 10000
OIT3 LV301 K22E LAC2 KAIN HEMO ALBU APOF A2MG
Protein [ ] (a.u.)
PSC
Early PSC-CCA
B
Non-PSC CCA biomarkers
Healthy
CCA
Early
Advanced
Early CCA vs Healthy
Protein
AUC
SEN (%)
SPE (%)
PPV (%)
NPV (%)
AI (%)
OR (95% CI)
SAMP 0.795 86.7 70.7 52.0 93.6 75.0 15.7 (3.1 – 80.5)
COF1 0.852 100.0 58.5 46.9 100.0 69.6 11.3 (2.9 – 45.1)
ACTB 0.784 93.3 56.1 43.8 95.8 66.1 8.1 (1.9 – 34.4)
ACTC 0.738 46.7 92.7 70.0 82.6 80.4 8.1 (1.9 – 34.4)
ADA10 0.754 60.0 87.8 64.3 85.7 80.4 10.8 (2.7 – 43.5)
TSN9 0.691 60.0 82.9 56.3 85.0 76.8 7.3 (2.0 – 27.1)
1 10 100 1000 10000
SAMP COF1 ACTB ACTC ADA10 TSN9
Protein [ ] (a.u.)
Healthy
Early CCA
C
Pan-CCA biomarkers
Healthy
PSC
Pan-CCA
Early
Advanced
Early Pan-CCA vs [Healthy + PSC]
Protein
AUC
SEN (%)
SPE (%)
PPV (%)
NPV (%)
AI (%)
OR (95% CI)
PIGR 0.770 90.0 58.8 35.3 95.9 65.0 12.8 (2.8 – 59.0)
FIBG 0.733 80.0 63.8 35.6 92.7 67.0 7.0 (2.1 – 23.0)
PPBT 0.722 65.0 73.8 38.2 89.4 72.0 5.2 (1.8 – 14.8)
A1AT 0.718 65.0 76.3 40.6 89.7 74.0 6.0 (2.1 – 17.1)
FIBB 0.671 55.0 78.8 39.3 87.5 74.0 4.5 (1.6 – 12.7)
RELN 0.706 80.0 61.3 34.0 92.5 65.0 6.3 (1.9 – 20.7)
VWF 0.701 70.0 67.5 35.0 90.0 68.0 5.9 (1.9 – 17.9)
IGHG4 0.734 75.0 66.3 35.7 91.4 68.0 5.2 (1.8 – 14.6)
LRP1 0.676 60.0 77.5 40.0 88.6 74.0 4.0 (1.2 – 13.0)
FGL1 0.674 80.0 50.0 28.6 90.9 56.0 9.9 (3.0 – 32.8)
APOA1 0.806 80.0 71.3 41.0 93.4 73.0 5.4 (1.7 – 17.6)
HPTR 0.721 80.0 57.5 32.0 92.0 62.0 9.6 (3.2 – 28.6)
APOL1 0.786 65.0 83.8 50.0 90.5 80.0 10.0 (3.2 – 28.6)
KLKB1 0.756 85.0 63.8 37.0 94.4 68.0 10.0 (2.7 – 36.9)
GELS 0.660 80.0 63.8 35.6 92.7 67.0 7.0 (2.1 – 23.0)
FLNA 0.685 80.0 56.3 31.4 91.8 61.0 5.1 (1.6 – 16.8)
ACTN1 0.672 70.0 68.8 35.9 90.2 69.0 5.1 (1.8 – 14.9)
ITIH2 0.718 50.0 85.0 45.5 87.2 78.0 5.7 (1.9 – 16.5)
1433Z 0.649 65.0 71.3 36.1 89.1 70.0 4.6 (1.6 – 13.0)
TLN1 0.683 50.0 82.5 41.7 86.8 76.0 4.7 (1.7 – 13.5)
VTDB 0.764 55.0 88.8 55.0 88.8 82.0 9.6 (3.1 – 29.6)
IMA8 0.677 95.0 36.3 27.1 96.7 48.0 10.8 (1.4 – 84.9)
1 10 100 1000 10000
IGHG4 APOA1 APOL1 ITIH2 VTDB
Protein [ ] (a.u.)
Healthy
PSC
Early Pan-CCA
D
Predictive biomarkers of CCA development in PSC
PSC-CCA biomarkers
HEMO ALBU
Pan-CCA biomarkers
FIBG FIBB FGL1 IGHA1 FLNA TLN1 IMA8
Protein [ ] (a.u.)
PSC
PSC to CCA

A **Early PSC-CCA biomarkers (specific)**

Protein	Early PSC-CCA *vs* Healthy						Early PSC-CCA *vs* early CCA						Ealy PSC-CCA *vs* [Healthy + PSC + early CCA]					
	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)
OIT3	0.932	100.0	85.4	45.5	100.0	87.0	0.960	100.0	93.3	83.3	100.0	95.0	0.901	100.0	80.0	20.8	100.0	81.0
LV301	0.902	100.0	73.2	31.3	100.0	76.1	0.840	100.0	80.0	62.5	100.0	85.0	0.840	100.0	71.6	15.6	100.0	73.0
K22E	0.800	100.0	70.7	29.4	100.0	73.9	0.960	100.0	93.3	83.3	100.0	95.0	0.865	100.0	80.0	20.8	100.0	81.0
LAC2	0.951	100.0	92.7	62.5	100.0	93.5	0.827	100.0	80.0	62.5	100.0	85.0	0.897	100.0	87.4	29.4	100.0	88.0
IGLL5	0.917	100.0	80.5	38.5	100.0	82.6	0.987	100.0	93.3	83.3	100.0	95.0	0.874	100.0	76.8	18.5	100.0	78.0
KAIN	0.883	100.0	70.7	29.4	100.0	73.9	0.987	100.0	93.3	83.3	100.0	95.0	0.899	100.0	71.6	15.6	100.0	73.0
HEMO	0.741	80.0	65.9	22.2	96.4	67.4	0.813	80.0	86.7	66.7	92.9	85.0	0.785	80.0	74.7	14.3	98.6	75.0
ALBU	0.868	100.0	70.7	29.4	100.0	73.9	0.693	60.0	80.0	50.0	85.7	75.0	0.815	100.0	55.8	10.6	100.0	58.0
APOF	0.946	80.0	97.6	80.0	97.6	95.7	0.867	80.0	93.3	80.0	93.3	90.0	0.897	80.0	90.5	30.8	98.9	90.0
A2MG	0.834	100.0	63.4	25.0	100.0	67.4	0.960	80.0	100.0	100.0	93.8	95.0	0.855	100.0	68.4	14.3	100.0	70.0

B **Early Non-PSC CCA biomarkers**

Protein	Early CCA *vs* PSC						Early CCA *vs* early PSC-CCA						Early CCA *vs* [Healthy + PSC + early PSC-CCA]					
	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)
SAMP	0.704	86.7	53.9	41.9	91.3	63.0	0.760	80.0	80.0	92.3	57.1	80.0	0.751	86.7	62.4	28.9	96.4	66.0
COF1	0.683	100.0	41.0	39.5	100.0	57.4	0.920	100.0	80.0	93.8	100.0	95.0	0.778	100.0	51.8	26.8	100.0	59.0
ACTB	0.607	33.3	92.3	62.5	78.3	75.9	0.720	93.3	60.0	87.5	75.0	85.0	0.699	86.7	48.2	22.8	95.4	54.0
ACTC	0.527	40.0	79.5	42.9	77.5	68.5	0.627	33.3	100.0	100.0	33.3	50.0	0.635	40.0	87.1	35.3	89.2	80.0
ADA10	0.675	60.0	79.5	52.9	83.8	74.1	0.893	80.0	100.0	100.0	62.5	85.0	0.726	60.0	84.7	40.9	92.3	81.0
TSN9	0.670	60.0	74.4	47.4	82.9	70.4	0.720	60.0	100.0	100.0	45.5	70.0	0.683	60.0	80.0	34.6	91.9	77.0

C **Early Pan-CCA biomarkers**

Protein	Early Pan-CCA *vs* Healthy						Early Pan-CCA *vs* PSC					
	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)	AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)
PIGR	0.952	90.0	100.0	100.0	95.4	96.7	0.578	65.0	59.0	44.8	76.7	61.0
FIBG	0.768	85.0	65.9	54.8	90.0	72.1	0.696	80.0	59.0	50.0	85.2	66.1
PPBT	0.831	85.0	78.1	65.4	91.4	80.3	0.606	50.0	74.4	50.0	74.4	66.1
A1AT	0.756	90.0	58.5	51.4	92.3	68.9	0.678	65.0	76.9	59.1	81.1	72.9
FIBB	0.724	60.0	78.1	57.1	80.0	72.1	0.615	55.0	74.4	52.4	76.3	67.8
RELN	0.729	60.0	82.9	63.2	81.0	75.4	0.682	80.0	61.5	51.6	85.7	67.8
VWF	0.763	70.0	78.1	60.9	84.2	75.4	0.636	55.0	74.4	52.4	76.3	67.8
IGHG4	0.784	75.0	75.6	60.0	86.1	75.4	0.683	75.0	56.4	46.9	81.5	62.7
LRP1	0.656	65.0	68.3	50.0	80.0	67.2	0.696	60.0	82.1	63.2	80.0	74.6
FGL1	0.667	95.0	36.6	42.2	93.8	55.7	0.681	80.0	56.4	48.5	84.6	64.4
APOA1	0.816	60.0	92.7	80.0	82.6	82.0	0.796	80.0	74.4	61.5	87.9	76.3
HPTR	0.683	50.0	82.9	58.8	77.3	72.1	0.760	85.0	69.2	58.6	90.0	74.6
APOL1	0.788	65.0	90.2	76.5	84.1	82.0	0.783	90.0	66.7	58.1	92.9	74.6
KLKB1	0.828	75.0	85.4	71.4	87.5	82.0	0.679	85.0	53.9	48.6	87.5	64.4
GELS	0.767	80.0	82.9	69.6	89.5	82.0	0.547	80.0	43.6	42.1	81.0	55.9
FLNA	0.768	80.0	70.7	57.1	87.9	73.8	0.597	80.0	41.0	41.0	80.0	54.2
ACTN1	0.718	75.0	70.7	55.6	85.3	72.1	0.623	70.0	61.5	48.3	80.0	64.4
ITIH2	0.760	50.0	92.7	76.9	79.2	78.7	0.674	90.0	43.6	45.0	89.5	59.3
1433Z	0.651	65.0	73.2	54.2	81.1	70.5	0.646	65.0	69.2	52.0	79.4	67.8
TLN1	0.734	75.0	65.9	51.7	84.4	68.9	0.628	50.0	74.4	50.0	74.4	66.1
VTDB	0.765	55.0	90.2	73.3	80.4	78.7	0.764	50.0	92.3	76.9	78.3	78.0
IMA8	0.690	95.0	39.0	43.2	94.1	57.4	0.663	85.0	43.6	43.6	85.0	57.6

D **Predictive biomarkers of CCA development in PSC**

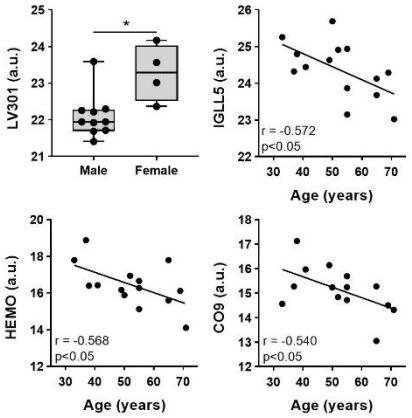
	Protein	PSC to CCA *vs* PSC						
		AUC	SEN (%)	SPE (%)	PPV (%)	NPV (%)	AI (%)	OR (95% CI)
PSC-CCA biomarkers	HEMO	0.705	60.0	79.5	42.9	88.6	75.5	5.8 (1.3 – 25.7)
	ALBU	0.726	100.0	48.7	33.3	100.0	59.2	20.0 (1.1 – 364.5)
Pan-CCA biomarkers	FIBG	0.759	90.0	64.1	39.1	96.2	69.4	16.1 (1.8 – 140.4)
	FIBB	0.728	90.0	56.4	34.6	95.7	63.3	11.6 (1.3 – 101.1)
	FGL1	0.776	80.0	79.5	50.0	93.9	79.6	15.5 (2.7 – 87.7)
	IGHA1	0.787	90.0	69.2	42.9	96.4	73.5	20.3 (2.3 – 178.3)
	FLNA	0.762	60.0	87.2	54.6	89.5	81.6	10.2 (2.1 – 49.3)
	TLN1	0.710	70.0	76.9	43.8	90.9	75.5	7.8 (1.7 – 36.4)
	IMA8	0.699	90.0	61.5	37.5	96.0	67.4	14.4 (1.7 – 125.4)
	CA19-9	0.500	30.0	75.0	25.0	79.4	65.2	1.3 (0.27 – 6.0)

Figure 4.8. related to Figure 4.7. Diagnostic test evaluation of specific early PSC-CCA biomarkers (**A**), early non-PSC CCA biomarkers (**B**), early Pan-CCA biomarkers (**C**) and of predictor biomarkers of CCA in PSC (**D**).

We next evaluated potential associations of these candidate biomarkers with demographic features (age, gender) and/or CCA subtype. Among PSC-CCA biomarkers, only LV301 levels varied with sex, while IGLL5, HEMO and CO9 negatively correlated with age (**Figure 4.9**). Considering biomarkers for the diagnosis of CCA in non-PSC background, only ADA10 displayed an association with patients' gender. Among pan-CCA candidate biomarkers, increased APOL1 levels were found in women, PIGR and FLNA levels were higher in younger people, while FIBG was the only candidate biomarker that might be affected by CCA subtype. The levels of all the other individual candidate biomarkers were independent to gender, age and CCA subtype.

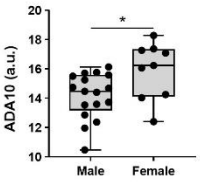
A **Specific PSC-CCA biomarkers**

Protein	Sex (male *vs* female)		Age (continuous)		CCA subtype (iCCA *vs* pCCA *vs* dCCA)	
	Student, t	p-value	Pearson, r	p-value	ANOVA, F	p-value
OIT3	-0.18	>0.05	0.33	>0.05	0.66	>0.05
LV301	-3.10	**<0.01**	-0.01	>0.05	1.65	>0.05
K22E	0.15	>0.05	-0.12	>0.05	0.92	>0.05
HV308	-0.08	>0.05	-0.35	>0.05	1.34	>0.05
LAC2	1.50	>0.05	-0.33	>0.05	1.63	>0.05
TFR1	0.20	>0.05	-0.37	>0.05	0.37	>0.05
IGLL5	0.07	>0.05	-0.57	**<0.05**	0.41	>0.05
PLCH1	-0.61	>0.05	-0.29	>0.05	0.64	>0.05
KAIN	0.72	>0.05	-0.41	>0.05	0.06	>0.05
HEMO	-0.68	>0.05	-0.57	**<0.05**	0.52	>0.05
ALBU	-0.26	>0.05	-0.18	>0.05	3.22	>0.05
APOF	0.28	>0.05	-0.50	>0.05	0.23	>0.05
CO9	0.68	>0.05	-0.54	**<0.05**	2.00	>0.05
A2MG	1.72	>0.05	0.37	>0.05	0.16	>0.05



B **Non-PSC CCA biomarkers**

Protein	Sex (male *vs* female)		Age (continuous)		CCA subtype (iCCA *vs* pCCA *vs* dCCA)	
	Student, t	p-value	Pearson, r	p-value	ANOVA, F	p-value
SAMP	0.06	>0.05	0.29	>0.05	0.64	>0.05
COF1	0.11	>0.05	0.09	>0.05	0.15	>0.05
ACTB	1.24	>0.05	-0.03	>0.05	0.20	>0.05
ACTC	-0.02	>0.05	-0.01	>0.05	0.07	>0.05
ADA10	-2.47	**<0.05**	-0.05	>0.05	0.02	>0.05
TSN9	-0.31	>0.05	-0.21	>0.05	0.92	>0.05



C **Pan-CCA biomarkers**

Protein	Sex (male *vs* female)		Age (continuous)		CCA subtype (iCCA *vs* pCCA *vs* dCCA)	
	Student, t	p-value	Pearson, r	p-value	ANOVA, F	p-value
PIGR	-0.96	>0.05	-0.37	**<0.05**	2.34	>0.05
FIBG	-0.15	>0.05	0.13	>0.05	4.48	**<0.05**
PPBT	-1.28	>0.05	-0.01	>0.05	1.47	>0.05
A1AT	-0.31	>0.05	0.09	>0.05	1.12	>0.05
FIBB	0.83	>0.05	0.09	>0.05	2.25	>0.05
RELN	-0.30	>0.05	-0.25	>0.05	1.48	>0.05
VWF	-0.33	>0.05	0.23	>0.05	0.44	>0.05
IGHG4	0.84	>0.05	0.13	>0.05	0.20	>0.05
LRP1	-0.48	>0.05	0.00	>0.05	2.68	>0.05
FGL1	0.67	>0.05	0.07	>0.05	1.67	>0.05
IGHA1	-1.01	>0.05	-0.15	>0.05	0.70	>0.05
APOA1	-1.49	>0.05	0.05	>0.05	1.16	>0.05
HPTR	-0.69	>0.05	-0.13	>0.05	0.03	>0.05
APOL1	-2.23	**<0.05**	-0.20	>0.05	0.92	>0.05
KLKB1	-0.02	>0.05	0.27	>0.05	0.33	>0.05
GELS	-0.39	>0.05	0.04	>0.05	0.51	>0.05
FLNA	-0.11	>0.05	-0.33	**<0.05**	0.18	>0.05
ACTN1	-0.14	>0.05	0.24	>0.05	0.40	>0.05
ITIH2	-1.01	>0.05	-0.04	>0.05	0.93	>0.05
1433Z	0.04	>0.05	-0.04	>0.05	0.23	>0.05
TLN1	-0.41	>0.05	-0.20	>0.05	0.23	>0.05
VTDB	-0.61	>0.05	0.12	>0.05	1.06	>0.05
IMA8	0.42	>0.05	-0.04	>0.05	0.19	>0.05

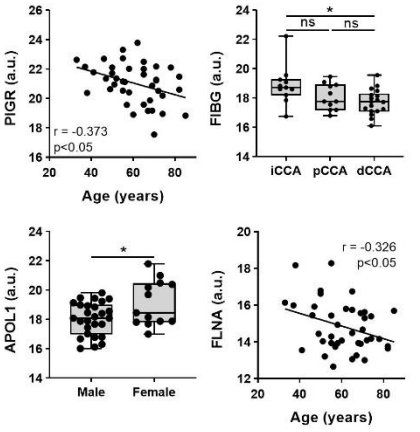


Figure 4.9. related to Figure 4.5. Association of the protein abundance in serum EVs with biological sex, age and CCA anatomical subtype in patients with concomitant PSC-CCA (**A**), in patients with CCA without PSC (**B**) and in patients with CCA regardless etiology (**C**).

In order to generate the best algorithms for the diagnosis of CCA in patients with or without PSC, only EV-proteins with levels invariable with sex, age and CCA subtype were selected for machine learning logit modelling. A binary logistic regression model combining 8 protein biomarkers (HV308/KAIN/HEMO/A2MG/VWF/APOA1/TLN1/IMA8) provided the maximum accuracy to distinguish patients with concomitant PSC-CCA compared to patients with isolated PSC, maintaining the optimal diagnostic power also at early-stage CCA (**Figure 4.10A**), with a 100% of accuracy. Noteworthy, the combination of PLCH1/FGL1 allowed the differential diagnosis of PSC-CCA vs PSC with high accuracy, even at early tumor stages, overpowering serum CA19-9 values (AUC=0.903 vs 0.608, respectively); moreover, patients with elevation of these 2 EV-protein biomarkers display 33.3-fold increased risk of CCA occurrence. When the absence of PSC is clinically confirmed, a logit model combining 5 biomarkers (SAMP/ACTC/RELN/APOA1/KLKB1) was 100% sensitive and specific for discriminating among patients with CCA compared to healthy individuals, also at early tumor stages (**Figure 4.10B**). A model combining only SAMP/A1AT might also help in the diagnosis of CCAs arising from non-PSC etiologies, providing an AUC value of 0.899 when compared with healthy individuals and 0.863 for the early diagnosis of CCA.

Considering the independent predictors of CCA development in patients with PSC (ALBU, FIBB, FGL1, IGHA1, TLN1 and IMA8), 100% of patients with PSC who displayed serum alterations in these 6 biomarkers developed CCA during follow-up while 100% of patients with PSC who displayed alterations in ≤3 of these biomarkers did not present biliary malignancy over time (**Figure 4.10C**).

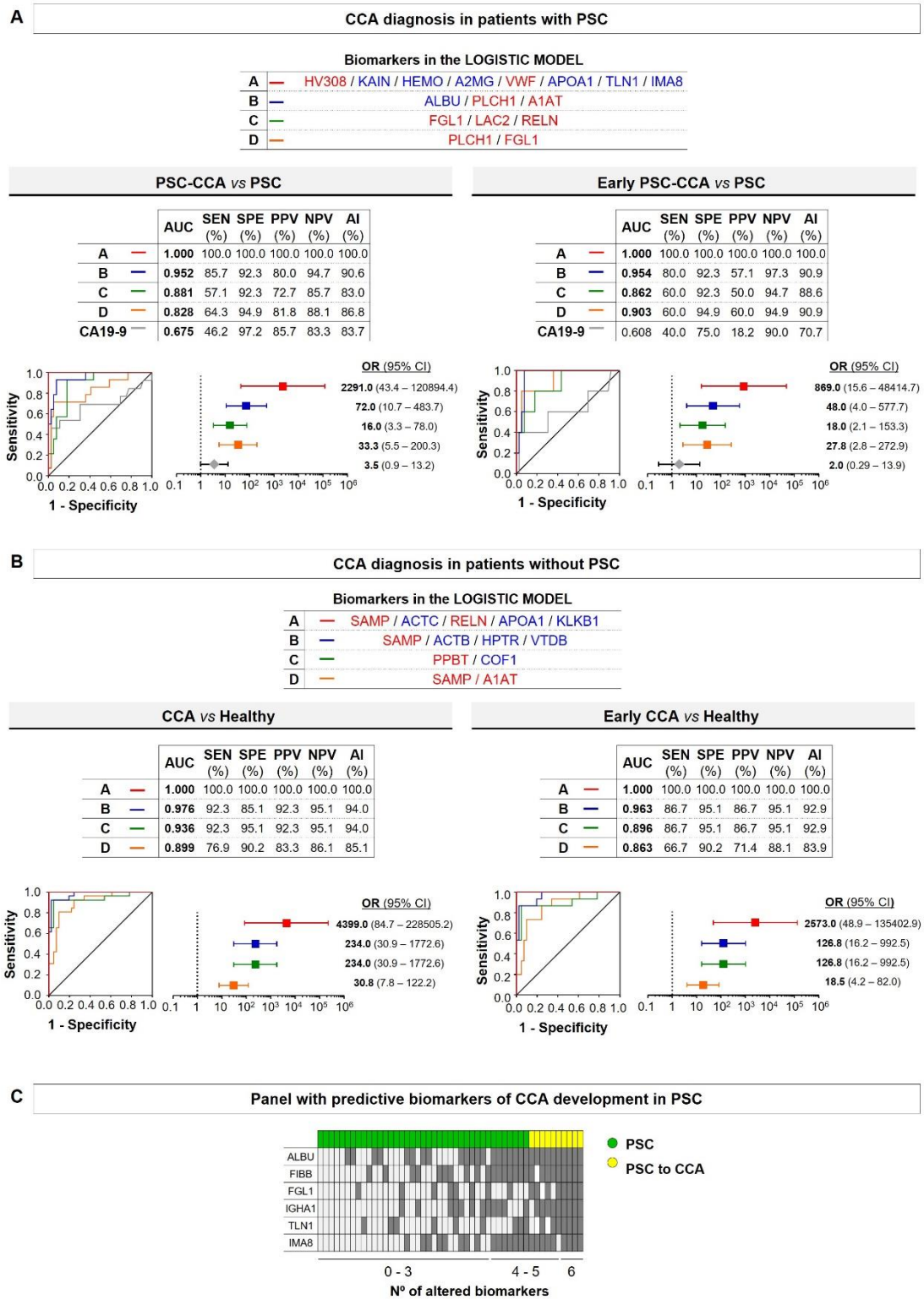


Figure 4.10. Combination of EV-protein biomarkers enables the accurate diagnosis of CCA in patients with or without PSC. Binary logistic regression models for (**A**) CCA diagnosis in patients with PSC, combining two to eight specific PSC-CCA and pan-CCA biomarkers which had individually higher (red font) or lower (blue font) abundance in serum EVs from patients with concomitant PSC-CCA, and for (**B**) CCA diagnosis in patients without PSC, combining two to five specific non-PSC CCA and pan-CCA biomarkers that had individually higher (red font) or lower (blue font) abundance in serum EVs from patients with CCA of non-PSC etiologies. (**C**) Panel of predictive CCA biomarkers with the proportion of PSC to CCA and isolated PSC properly identified according to the number of altered biomarkers.

3.4 Candidate EV-protein biomarkers are expressed mainly in normal liver and differ across specific liver cell populations

In order to decipher the potential origin of the candidate serum EV-protein biomarkers for CCA, their gene expression was first analyzed in a human multi-organ (n=61 tissues/organs) transcriptomic dataset (**Figure 4.11A, Figure 4.12**). Data revealed that almost all these EV-protein biomarkers are expressed in the liver and/or the gallbladder. Interestingly, some candidate EV-biomarkers including HEMO (HPX), ALBU (ALB), APOF (APOF), CO9 (C9), SAMP (APCS), FIBG (FGG), A1AT (SERPINA1), FIBB (FGB), FGL1 (FGL1), HPTR (HPR), KLKB1 (KLKB1), ITIH2 (ITIH2) and VTDB (GC) were almost exclusively expressed in hepatobiliary tissues, representing more than 80% of the total expression detected in all human organs (**Figure 4.11A**).

Human liver scRNA-seq showed that the expression of some candidate biomarkers was cell type-specific while others were indiscriminately expressed along different liver cell types (**Figure 4.11B; Figure 4.13A**). For instance, PLCH1 (PLCH1) and PIGR (PIGR) were almost exclusively expressed in cholangiocytes; endothelial cells had the highest expression of OIT3 and VWF; the main source of LV301 (IGLV3-1) and IGLL5 (IGLL5) was B cells and hepatocytes highly expressed APOF, APOL1 (APOL1) and GELS (GSN).

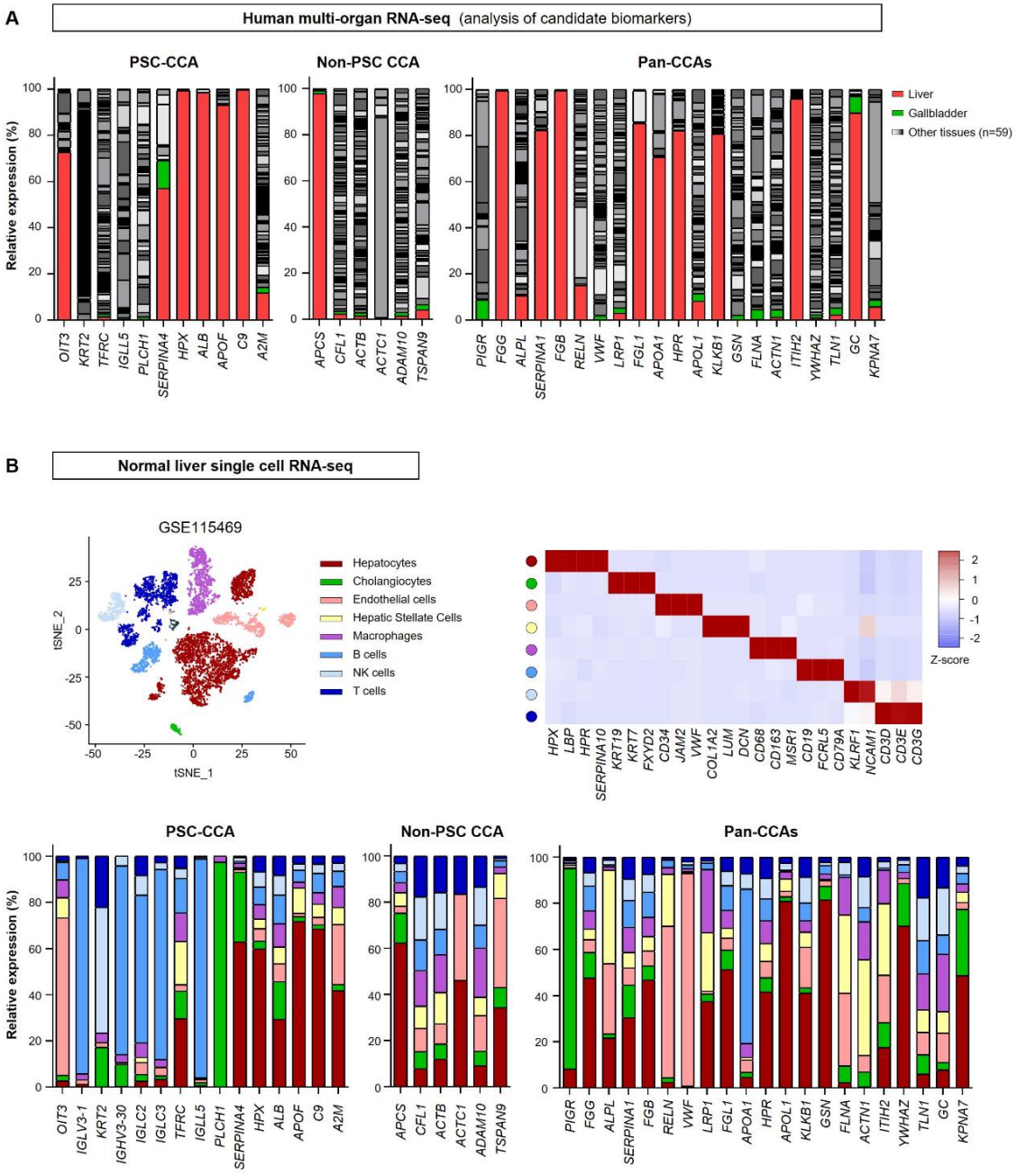


Figure 4.11. Human multi-organ transcriptomic studies and single-cell RNA-seq datasets reveals the potential tissue and cell type of origin of serum EV-protein biomarkers. (**A**) Normalized relative expression of candidate serum biomarkers in 61 human tissues/organs collected in the Consensus dataset of the Human Protein Atlas. (**B**) Normal liver scRNA-seq from the GSE115469 source classified according to indicated cell markers revealed the normalized relative expression of candidate biomarkers in each liver cell type.

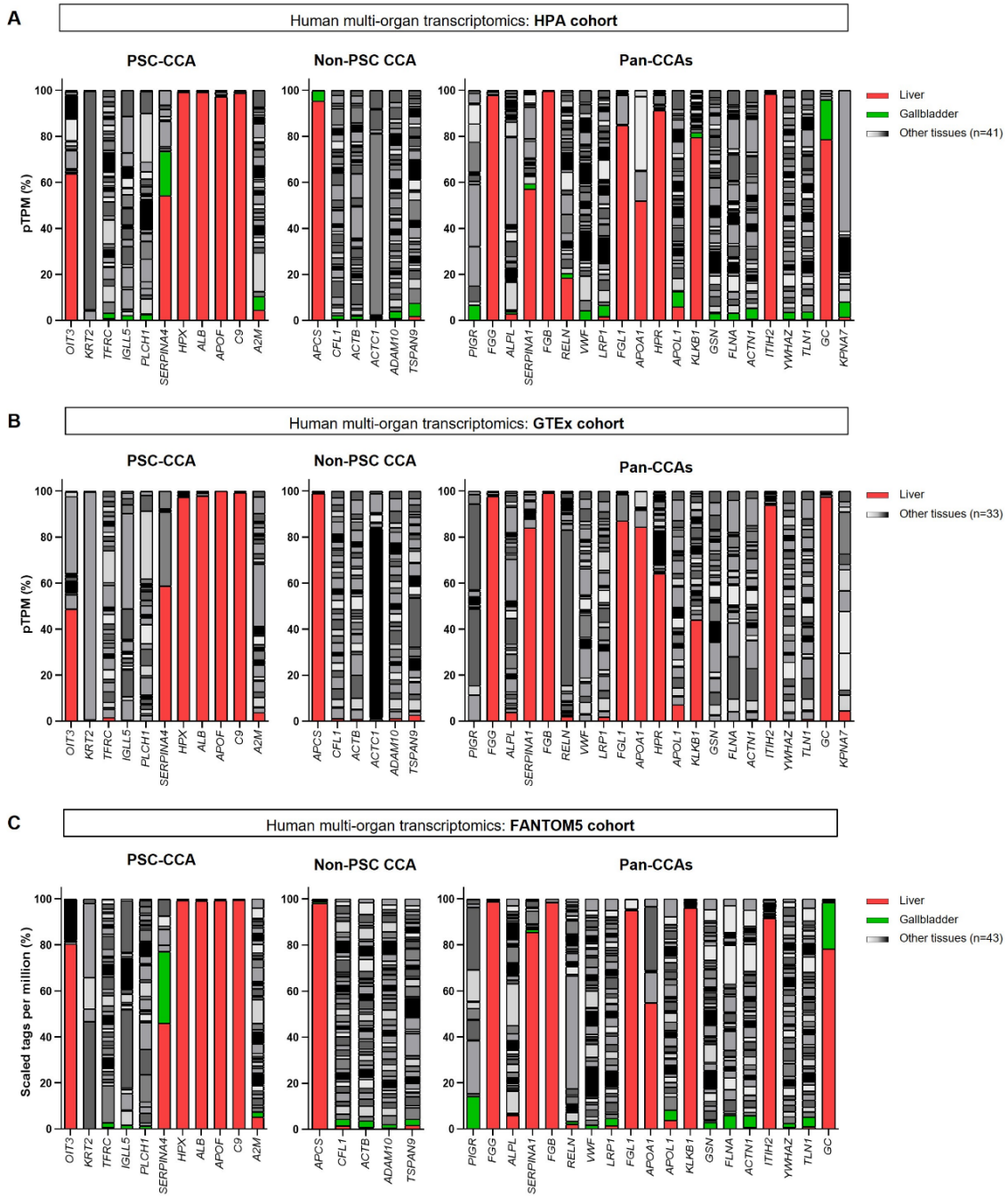


Figure 4.12: related to Figure 4.11A. Analysis of candidate serum biomarkers in the three tissue cohorts which are summarized in the Consensus gene data of The Human Protein Atlas: HPA (**A**), GTEx: Genotype-Tissue Expression (**B**) and FANTOM5: Functional Annotation of Mammalian Genomes 5 (**C**).

3.5 Single-cell RNA-seq data reveals specific cell types within human CCA tumors expressing EV-protein biomarkers

Considering that scRNA-seq studies published to date included only tumor samples from patients with non-PSC CCA, we next investigated the cell populations within human CCA tumors expressing the aforementioned pan-CCA or non-PSC CCA biomarkers (**Figure 4.14, Figures 4.13B and 4.15**). When analyzing the single-cell transcriptome of GSE151530 and GSE125449 datasets, composed by CCA tumor samples from 12 and 10 patients, respectively, the highest percentage of cells expressing PIGR, FGG, FGB, RELN, FGL1 and APCS transcripts were malignant cholangiocytes, while VWF- and ALPL-expressing cells were principally endothelial cells. LRP1-positive cells constituted mostly malignant cholangiocytes, hepatic stellate cells and macrophages, and the expression of SERPINA1 was more widespread, being detected in all the main tumor cells with a predominant increase in tumor cholangiocytes and macrophages.

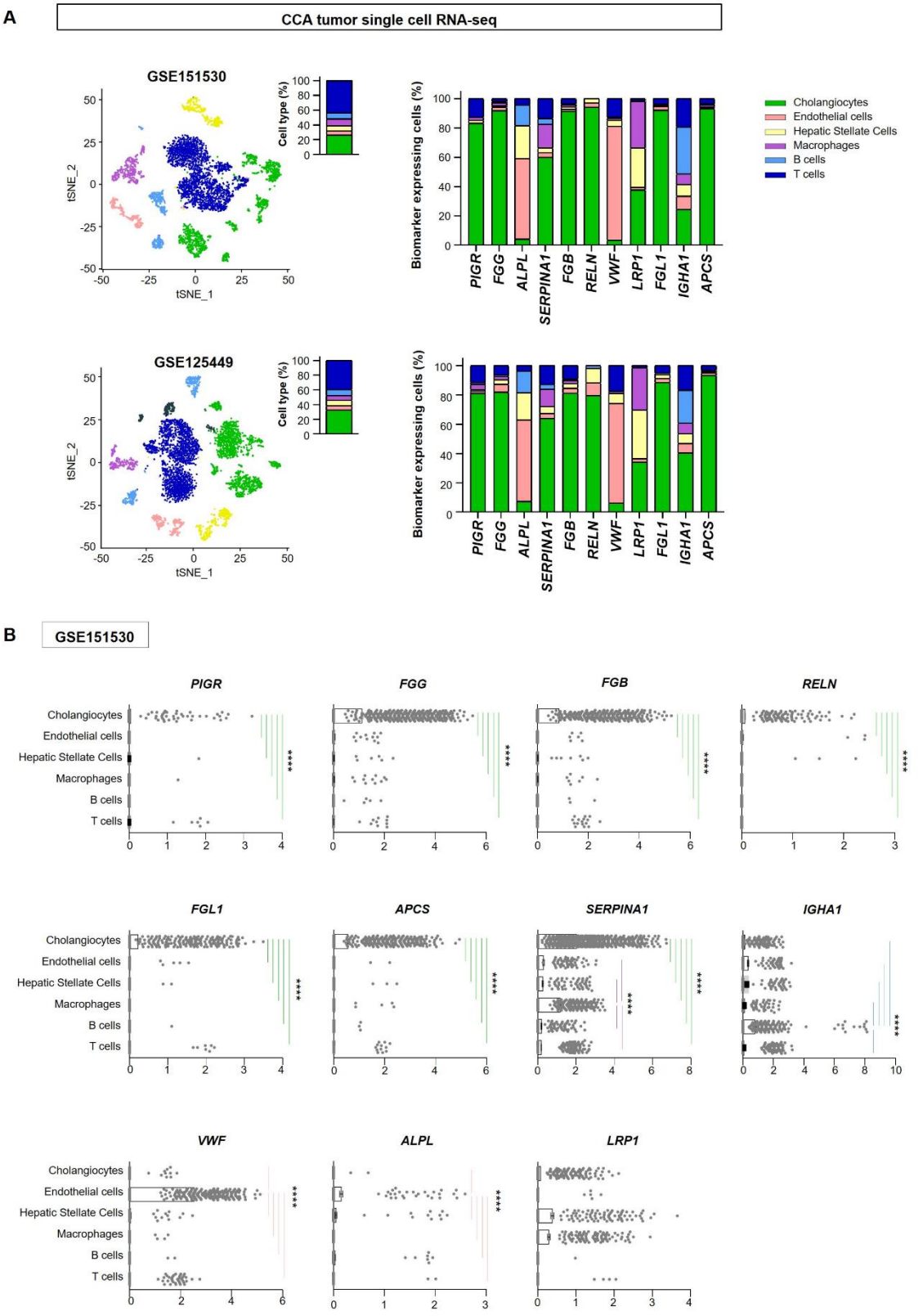


Figure 4.14. CCA tumor cell type expression of candidate serum biomarkers. (**A**) tSNE plots and proportion of cell types in two datasets comprising fresh iCCA tumor biopsies from 12 and 10 patients, respectively. Comparison of biomarker expressing cells in the tumor scRNA-seq analysis. (**B**) Relative biomarker expression within the single-cell types of CCA tumors. Non-parametric Kruskal-Wallis test was used, followed by a posteriori Dunns test. Data are shown as mean ±SEM. **** represent p-values of <0.0001.

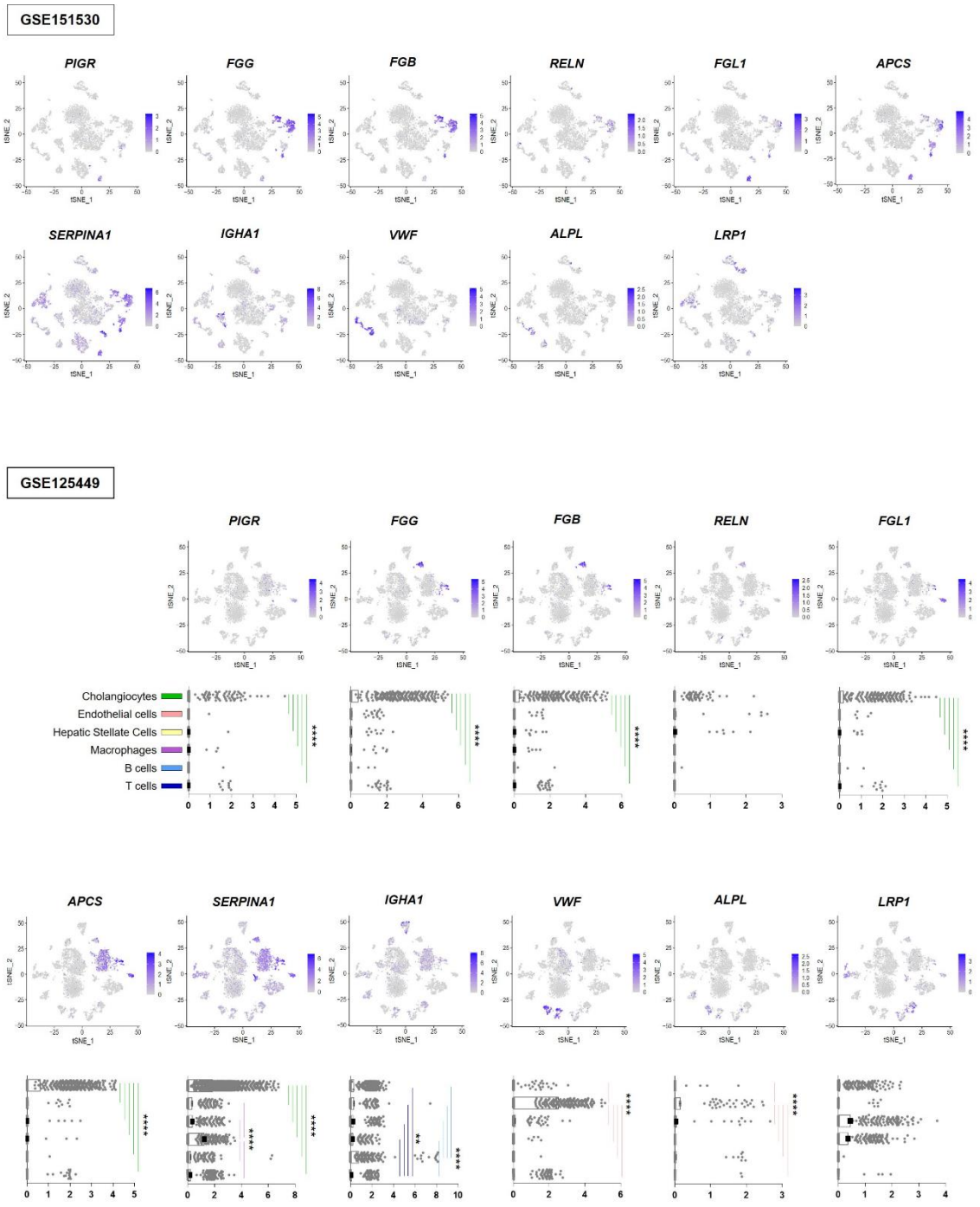


Figure 4.15. related to Figure 4.14. tSNE plots and normalized relative biomarker expression of CCA tumor-constituting single-cell types. Non-parametric Kruskal-Wallis test was used, followed by a posteriori Dunns test. Data are shown as mean ±SEM. **** represent p-values of <0.0001.

3.6 Serum EVs hold proteins with prognostic capacity for patients with CCA

Univariable analysis revealed that 27 EV-proteins may also predict prognosis in patients with CCA (**Figure 4.16**). Multivariable analysis, including each candidate biomarker as well as demographic and clinical features (age, sex, tumor stage, CCA subtype, presence of PSC, cirrhosis and serum levels of CA19-9, total bilirubin, ALP, ALT, AST and GGT), revealed that the levels of FCN2, MMRN1, SDPR, ACTC, ADIPO, CO4A, FA9, PSB3 and PSA2 were independent predictors of overall survival (OS) (**Figure 4.17A,B).** Furthermore, a panel comprised by FCN2, SDPR and FA9 showed the best survival predictive capacity (**Figure 4.17C**). Of note, the median OS (mOS) of patients with high levels of three biomarkers (based on the cut-off) was superior to 120 months. In contrast, lower levels of one of these biomarkers were associated with shorter mOS (33.5 months). Remarkably, the risk of death of patients with CCA presenting decreased levels of 2 or 3 of the mentioned biomarkers was increased by 30 times [HR=30.0 (6.6 – 136.1)] when compared with patients with high levels of these 3 proteins, with a mOS of 6.9 months.

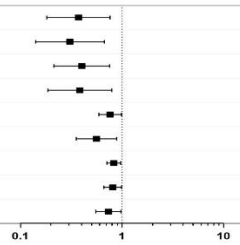
Clinical variables	Univariable analysis		
	HR	95% CI	p-Value
Age, (continuous)	0.994	0.963 – 1.027	>0.05
Sex, male (*vs* female)	1.017	0.421 – 2.460	>0.05
Tumor stage, III-IV (*vs* I-II)	5.451	2.136 – 13.909	**<0.001**
CCA subtype, (*vs* iCCA)	-	-	>0.05
pCCA	1.352	0.465 – 3.927	>0.05
dCCA	0.849	0.307 – 2.347	>0.05
PSC, Yes (*vs* No)	1.704	0.738 – 3.937	>0.05
CA19-9, (continuous)	1.00006	0.99997 – 1.0002	>0.05
Total bilirubin, (continuous)	1.004	1.001 – 1.007	**<0.05**
ALP, (continuous)	1.000	0.998 – 1.002	>0.05
ALT, (continuous)	0.998	0.994 – 1.001	>0.05
AST, (continuous)	0.998	0.995 – 1.002	>0.05
GGT, (continuous)	0.9999	0.9992 – 1.0005	>0.05
Cirrhosis, Yes (*vs* No)	0.044	0.000 – 107.643	>0.05

Protein variables	Univariable analysis		
	HR	95% CI	p-Value
C1S	2.479	1.001 – 6.138	**<0.05**
APOE	2.313	1.267 - 4.221	**<0.01**
LG3BP	1.878	1.062 - 3.324	**<0.05**
KV121	1.652	1.025 - 2.661	**<0.05**
HSP7C	1.546	1.035 - 2.311	**<0.05**
TFR1	1.492	1.067 - 2.086	**<0.05**
IGHA2	1.428	1.112 - 1.835	**<0.01**
APOC1	1.400	1.096 - 1.788	**<0.01**
KV306	1.351	1.027 - 1.779	**<0.05**
FCN1	1.328	1.021 - 1.729	**<0.05**
IMA8	1.272	1.006 - 1.609	**<0.05**
TTYH3	1.152	1.011 - 1.312	**<0.05**
PSA2	0.889	0.792 - 0.997	**<0.05**
MYL6B	0.888	0.804 - 0.981	**<0.05**
PSA3	0.866	0.761 - 0.985	**<0.05**
PSB3	0.864	0.779 - 0.958	**<0.01**
FA9	0.859	0.773 - 0.954	**<0.01**
CO4A	0.834	0.707 - 0.983	**<0.05**
ADIPO	0.765	0.635 - 0.922	**<0.01**
PLEK	0.746	0.570 - 0.976	**<0.05**
CO4B	0.712	0.520 - 0.975	**<0.05**
ACTC	0.700	0.533 - 0.920	**<0.05**
SDPR	0.635	0.461 - 0.874	**<0.01**
MMRN1	0.614	0.392 - 0.963	**<0.05**
GELS	0.587	0.368 - 0.935	**<0.05**
HEMO	0.568	0.364 - 0.885	**<0.05**
FCN2	0.404	0.246 - 0.662	**<0.001**

Figure 4.16: related to Figure 4.17. Univariable Cox regression analysis of the clinical parameters age, biological sex, tumor stage, anatomical CCA subtype, presence of PSC, presence of cirrhosis, values of biochemical parameters (*i.e.*, CA19-9, total bilirubin, ALP, ALT, AST and GGT) and of the proteins identified in serum EVs of patients with CCAs regardless etiology.

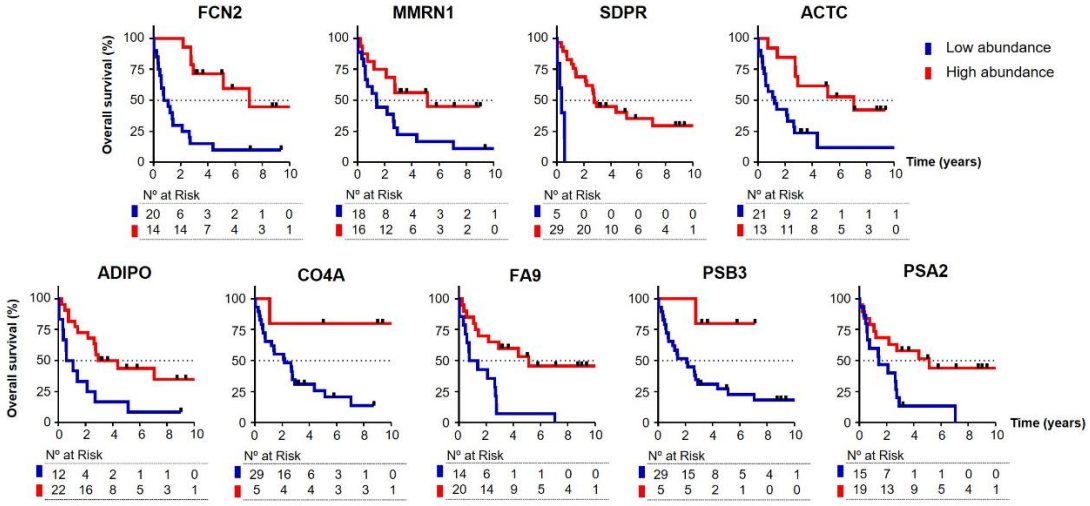
A

Multivariable analysis[a]	[a] multivariable analysis of each protein candidate with the following clinical parameters: **age** (continuous), **sex** (male *vs* female), **tumor stage** (III-IV *vs* I-II), **CCA subtype** (pCCA, dCCA *vs* iCCA), **PSC** (Yes *vs* No), **CA19-9** (continuous), **total bilirubin** (continuous), **ALP** (continuous), **ALT** (continuous), **AST** (continuous), **GGT** (continuous) and **cirrhosis** (Yes *vs* No)		
Protein	HR (95% CI)	1 / HR	p-Value
FCN2	**0.373** (0.182 – 0.764)	2.68	**<0.01**
MMRN1	**0.307** (0.141 – 0.669)	3.26	**<0.01**
SDPR	**0.402** (0.214 – 0.753)	2.49	**<0.01**
ACTC	**0.385** (0.187 – 0.795)	2.60	**<0.01**
ADIPO	**0.765** (0.593 – 0.988)	1.31	**<0.05**
CO4A	**0.560** (0.354 – 0.888)	1.79	**<0.05**
FA9	**0.831** (0.712 – 0.969)	1.20	**<0.05**
PSB3	**0.810** (0.663 – 0.991)	1.23	**<0.05**
PSA2	**0.735** (0.554 – 0.976)	1.36	**<0.05**



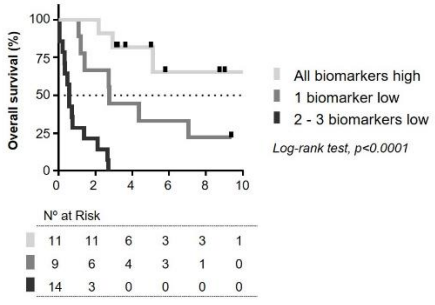
B

Categorized variables	Cox regression analysis			Log-rank (Mantel-Cox) test		
	HR (95% CI)	1 / HR	p-Value	**Median survival** (months) **[low abundance]**	**Median survival** (months) **[high abundance]**	p-Value
FCN2	**0.187** (0.072 – 0.484)	5.35	**<0.001**	**9.3** (0.1 – 18.5)	**85.8** (33.7 – 137.8)	**<0.001**
MMRN1	**0.388** (0.165 – 0.913)	2.58	**<0.05**	**17.0** (8.5 – 25.4)	**62.5** (0.0 – 133.9)	**<0.05**
SDPR	**0.060** (0.014 – 0.260)	16.67	**<0.001**	**4.4** (1.3 – 7.5)	**33.5** (28.7 – 38.3)	**<0.001**
ACTC	**0.291** (0.116 – 0.733)	3.44	**<0.01**	**14.8** (3.4 – 26.2)	**85.8** (10.9 – 160.7)	**<0.01**
ADIPO	**0.342** (0.151 – 0.774)	2.92	**<0.01**	**7.27** (0 – 17.9)	**35.6** (9.0 – 62.2)	**<0.01**
CO4A	**0.142** (0.019 – 1.066)	7.04	>0.05	**26.5** (0.22 – 52.8)	**> 120**	**<0.05**
FA9	**0.256** (0.110 – 0.597)	3.91	**<0.01**	**9.3** (0.0 – 24.0)	**62.5** (-)	**<0.001**
PSB3	**0.153** (0.021 – 1.138)	6.54	>0.05	**25.8** (9.0 – 42.6)	**> 120**	**<0.05**
PSA2	**0.342** (0.147 – 0.794)	2.92	**<0.05**	**17.2** (0.0 – 38.5)	**62.5** (12.4 – 112.7)	**<0.01**



C

FCN2 / SDPR / FA9



	All high	1 low	2 - 3 low
HR (95% CI)	1 (Ref.)	3.71 (0.95 – 14.5)	30.0 (6.6 – 136.1)
Median survival, *months* (95% CI)	> 120	33.5 (33.2 – 33.8)	6.9 (4.5 – 9.3)
Log-rank test, p-Value	-	<0.05 (*vs high*)	<0.0001 (*vs high*)
	-	-	<0.001 (*vs low in 1*)

Figure 4.17. Association of serum EV-protein levels with patients' outcome. (**A**) Multivariable analysis including the significant univariable EV-protein candidates and the clinical parameters of age, biological sex, tumor stage, CCA subtype, PSC, cirrhosis and the biochemical values of CA19-9, total bilirubin, ALP, ALT, AST and GGT. (**B**) Cox regression analysis and log-rank (Mantel-Cox) test of candidate prognostic biomarkers categorized according to the optimal cut-off value of each variable. Kaplan Meier curves of the overall survival for each prognostic biomarkers in a follow-up of 10 months. (**C**) Kaplan Meier curve, Cox regression analysis and log-rank test of patients with CCA according to the combinatory of FCN2-SDPR-FA9 biomarkers.

4. DISCUSSION

The precise, non-invasive, and early diagnosis of CCA remains a major challenge, particularly in patients at high risk, such as those with PSC. Here, we provide a high-throughput proteomic analysis of serum EVs from patients with PSC-related CCA, non-PSC CCAs, patients with PSC who developed CCA during follow-up, isolated PSC and healthy individuals, identifying novel biomarkers to predict CCA development in patients with PSC as well as for the early and accurate diagnosis of PSC-related CCA or CCAs regardless etiology. Proteins present in serum EVs also allowed the estimation of survival in patients with CCA. Human multi-organ transcriptomic analysis revealed that most diagnostic biomarkers were abundantly expressed in hepatobiliary tissues. scRNA-seq analysis of normal liver and CCA tumors pointed out that the expression of candidate biomarkers was cell-specific, being particularly expressed in malignant cholangiocytes within CCA tumors, thus reinforcing this novel tumor cell-derived liquid biopsy strategy (**Figure 4.18**).

Although the majority of CCAs emerge without a clear etiology, PSC is a well-established risk factor. Up to 15-20% of patients with PSC might develop CCA during their lifetime, which commonly appears within the first year after PSC diagnosis and in younger people (40-50 years), when compared with CCAs from other etiologies (~65 years).[33–35] In fact, CCA is currently responsible for more than 30% of PSC-associated deaths,[36] constituting a substantial health and social problem. Current screening strategies for CCA diagnosis in patients with PSC have scant clinical value due to their low accuracy. Serum CA19-9 levels are generally not elevated in early stages, are also increased in ~30% of patients with isolated PSC and up to 7% of the general population are unable to express CA19-9 due to FUT3 activity deficiency,[37] strongly limiting the usefulness of serum CA19-9 as a routine diagnostic screening tool. Moreover, at the radiological level, benign biliary strictures in patients with PSC closely resemble the initial malignant changes, which makes the early and appropriate diagnosis of CCA in patients with PSC extremely challenging. All these limitations markedly impact on CCA detection, which is an accidental event in up to 40% of PSC-CCA cases, when liver transplantation was required or even at autopsy.[33] Previous studies have proposed several serum biomarkers for the differential diagnosis of CCA and PSC, although most of them did not include patients with concomitant PSC-CCA,[37] raising major concerns when translating these findings into clinics.

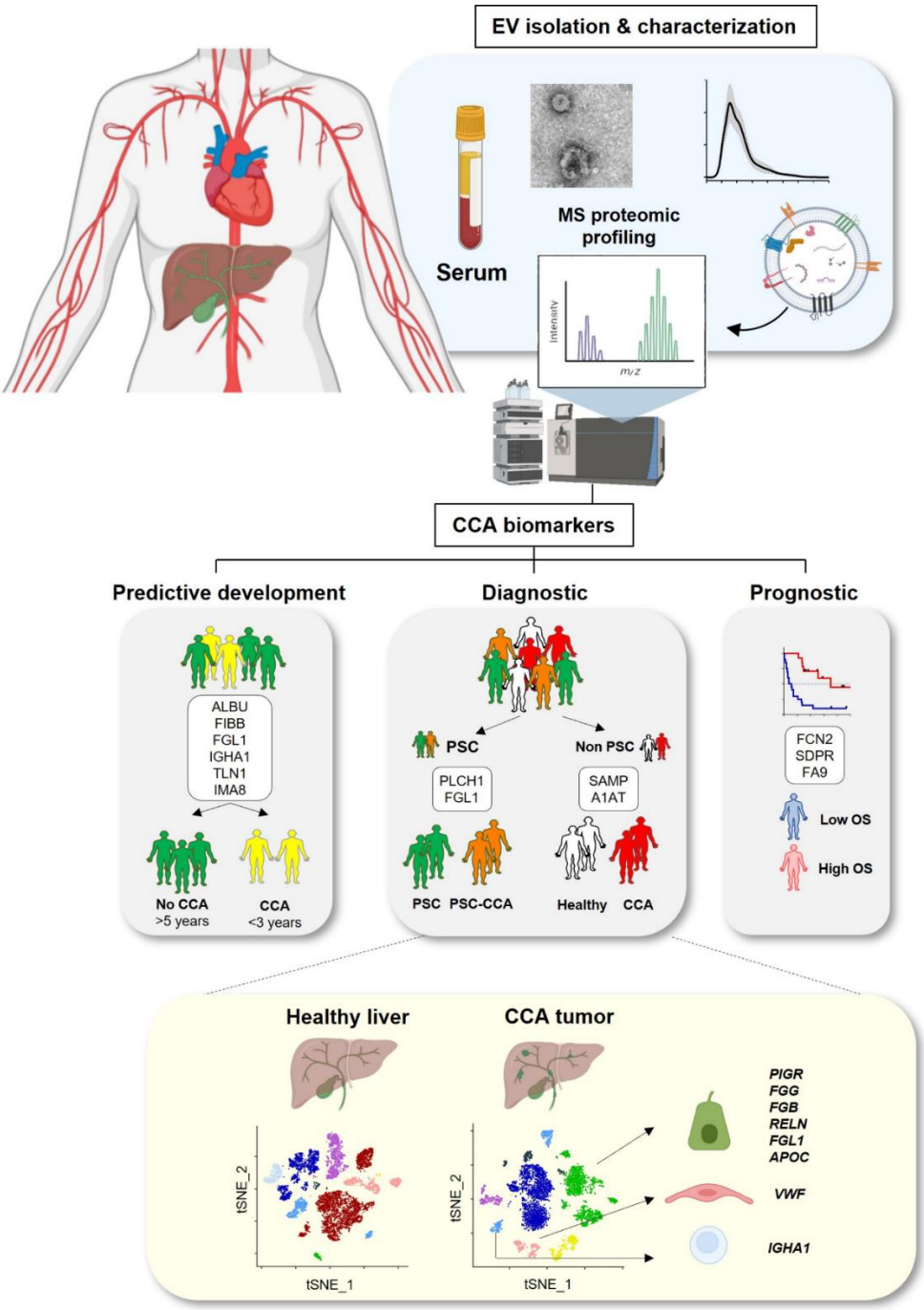


Figure 4.18. Predictive, diagnostic, and prognostic liquid biopsy biomarkers for CCA.

Our study contributed with novel protein biomarkers either specific for PSC-CCA, non-PSC CCA or general for all CCAs, which, individually or in combination, significantly improved the diagnostic capacity when compared to serum CA19-9 levels alone. Most of these novel biomarkers are independent of biological sex, age and anatomical CCA subtype, and revealed high diagnostic accuracy also at early tumor stages, discarding all these potential bias. Similarly, considering that >95% of patients with CCA included in our cohort do not present liver cirrhosis at sampling, the proposed candidate CCA

biomarkers are not associated with cirrhosis. Our data indicate that combination of PLCH1/FLG1/SAMP/A1AT might be particularly useful for the early diagnosis of PSC-CCA and CCA arising in patients without PSC, respectively. Furthermore, by using serum samples from patients with PSC that had no clinical evidence of malignant masses at the time of sampling but that developed CCA during follow-up, we proposed ALBU, FIBB, FLG1, IGHA1, TLN1 and IMA8 as novel predictive biomarkers of CCA development in patients with PSC, which were altered up to 3 years before the clinical evidence of tumor masses.

EVs were shown to recapitulate the phenotype and activities of their cells of origin,[38] mimicking the nature and behavior of source tissues. Considering that CCAs are highly heterogeneous, desmoplastic and stroma-enriched tumors,[2,3] scRNA-seq analysis might provide a better characterization of biomarker-expressing cell populations. The majority of the most promising biomarkers for CCA, which displayed an increased abundance in serum EVs when compared with controls, were found chiefly expressed in malignant cholangiocytes. Noteworthy, unravelling the type of cells that might be actively secreting biomarker-containing EVs could also postulate them as novel potential cell-specific therapeutic targets for CCA. For instance, PIGR, A1AT and FGL1, candidate biomarkers that were predominantly expressed in malignant cholangiocytes, have been studied in other cancer types and characterized as key contributors of tumorigenesis.[39–44] Consequently, their role in cholangiocarcinogenesis deserved future analysis.

Apart from harboring diagnostic and potential pathogenic role in CCA, we were also able to identify novel EV-protein prognostic biomarkers that may aid to predict OS in patients regardless disease etiology, tumor stage, serum CA19-9 levels, CCA subtype, cirrhosis and several other biochemical parameters. We herein propose a prognostic panel including FCN2, SDPR and FA9, where patients displaying decreased serum levels of at least 2 of these biomarkers presented 30-fold increased risk of death compared to having high levels in these 3 biomarkers. Although no information is available regarding the prognostic value of FA9 in cancer, lower FCN2 levels were included in a 6-gene panel to predict pathological complete response to neoadjuvant treatment in triple-negative breast cancer.[45] Similarly, decreased tumor expression of SDPR in renal cell carcinoma,[46] papillary thyroid cancer,[47] lung cancer[48] and HCC[49] was markedly associated with worse prognosis. Nevertheless, up to now, no reports have highlighted their potential value as circulating prognostic biomarkers.

This study has also limitations, primarily the number of patients with PSC-CCA included. Although displaying increased incidence trends, CCAs are rare tumors and patients with concomitant PSC-CCA only represent one fifth of them, resulting PSC-CCA sample

collection extremely challenging and samples very precious. In this regard, it is important to highlight that we only included strictly-selected patients, with biopsy-proven CCA confirmation, increasing the robustness of our analysis. Furthermore, by including patients with PSC that developed CCA but did not present any clinical evidence of tumor at the time of sample collection, we were able to identify predictive biomarkers of CCA. Although no clinical features of tumor development were observed in these patients, we may not discard that small, undetectable lesions might be already present at sampling. Consequently, these predictive biomarkers should be prospectively validated in a near future.

In conclusion, we here demonstrated the capacity of serum EVs to contain protein biomarkers for the prediction of CCA development as well as for the early tumor detection in individuals with PSC background, in individuals without PSC and also for CCAs regardless of disease etiology, overpowering the current gold standard serum CA19-9. This study reinforces the idea that CCAs arising from different etiologies (*e.g*., PSC vs non-PSC) may contain common and different serum EV-proteins, and therefore, there is a need to use proper, well-defined biomarkers to identify CCA in specific patient subgroups, moving a step forward into the personalized diagnosis of CCA. This, together with the fact that most of these candidate biomarkers are expressed in CCA tumors and preferentially in malignant cholangiocytes, reinforces our approach as an innovative tumor cell-derived liquid biopsy strategy. In order to determine the translational capacity of these novel predictive, diagnostic and prognostic biomarkers, a next validation phase using easily-transferable techniques, such as ELISA, will be used. This would open a new avenue for the early non-invasive diagnosis of CCA, consequently enabling a prompt therapeutic intervention and improving patients' welfare and outcome.

REFERENCES

1. Banales JM, Cardinale V, Carpino G, et al. Expert consensus document: Cholangiocarcinoma: current knowledge and future perspectives consensus statement from the European Network for the Study of Cholangiocarcinoma (ENS-CCA). *Nat Rev Gastroenterol Hepatol*. 2016;13(5):261-280. doi:10.1038/nrgastro.2016.51

2. Banales JM, Marin JJG, Lamarca A, et al. Cholangiocarcinoma 2020: the next horizon in mechanisms and management. *Nat Rev Gastroenterol Hepatol*. 2020;17(9):557-588. doi:10.1038/s41575-020-0310-z

3. Rodrigues PM, Olaizola P, Paiva NA, et al. Pathogenesis of Cholangiocarcinoma. *Annu Rev Pathol Mech Dis*. 2021;16:433-463. doi:10.1146/annurev-pathol-030220-020455

4. Razumilava N, Gores GJ. Cholangiocarcinoma. *Lancet*. 2014;383(9935):2168-2179. doi:10.1016/S0140-6736(13)61903-0

5. Bertuccio P, Malvezzi M, Carioli G, et al. Global trends in mortality from intrahepatic and extrahepatic cholangiocarcinoma. *J Hepatol*. 2019;71(1):104-114. doi:10.1016/j.jhep.2019.03.013

6. Macias RIR, Kornek M, Rodrigues PM, et al. Diagnostic and prognostic biomarkers in cholangiocarcinoma. *Liver Int*. 2019;39(S1):108-122. doi:10.1111/liv.14090

7. Valle JW, Borbath I, Khan SA, et al. Biliary cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2016;27(suppl 5):v28-v37. doi:10.1093/annonc/mdw324

8. Clements O, Eliahoo J, Kim JU, Taylor-Robinson SD, Khan SA. Risk factors for intrahepatic and extrahepatic cholangiocarcinoma: A systematic review and meta-analysis. *J Hepatol*. 2019;72(1). doi:10.1016/j.jhep.2019.09.007

9. Karlsen TH, Folseraas T, Thorburn D, Vesterhus M. Primary sclerosing cholangitis – a comprehensive review. *J Hepatol*. 2017;67(6):1298-1323. doi:10.1016/j.jhep.2017.07.022

10. Jiang X, Karlsen TH. Genetics of primary sclerosing cholangitis and pathophysiological implications. *Nat Rev Gastroenterol Hepatol*. 2017;14(5):279-295. doi:10.1038/nrgastro.2016.154

11. Lapitz A, Arbelaiz A, Olaizola P, et al. Extracellular vesicles in hepatobiliary malignancies. *Front Immunol*. 2018;9(OCT). doi:10.3389/fimmu.2018.02270

12. Thietart S, Rautou PE. Extracellular vesicles as biomarkers in liver diseases: A clinician's point of view. *J Hepatol*. 2020;73(6):1507-1525. doi:10.1016/j.jhep.2020.07.014

13. Yáñez-Mó M, Siljander PRM, Andreu Z, et al. Biological properties of extracellular vesicles and their physiological functions. *J Extracell Vesicles*. 2015;4(2015):1-60. doi:10.3402/jev.v4.27066

14. Raposo G, Stoorvogel W. Extracellular vesicles: Exosomes, microvesicles, and friends. *J Cell Biol*. 2013;200(4):373-383. doi:10.1083/jcb.201211138

15. Hirsova P, Ibrahim SH, Verma VK, et al. Extracellular vesicles in liver pathobiology: Small particles with big impact. *Hepatology*. 2016;64(6):2219-2233. doi:10.1002/hep.28814

16. Masyuk AI, Huang BQ, Ward CJ, et al. Biliary exosomes influence cholangiocyte regulatory mechanisms and proliferation through interaction with primary cilia. *Am J Physiol - Gastrointest Liver Physiol*. 2010;299(4). doi:10.1152/ajpgi.00093.2010

17. González E, Falcón-Pérez JM. Cell-derived extracellular vesicles as a platform to identify low-invasive disease biomarkers. *Expert Rev Mol Diagn*. 2015;15(7):907-923. doi:10.1586/14737159.2015.1043272

18. Arbelaiz A, Azkargorta M, Krawczyk M, et al. Serum extracellular vesicles contain protein biomarkers for primary sclerosing cholangitis and cholangiocarcinoma. *Hepatology*. 2017;66(4):1125-1143. doi:10.1002/hep.29291

19. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Management of cholestatic liver diseases. *J Hepatol*. 2009;51(2):237-267. doi:10.1016/j.jhep.2009.04.009

20. Edge SB, Byrd DR, Fritz AG, et al. *AJCC Cancer Staging Manual*. 7th editio. Springer; 2010.

21. Théry C, Witwer KW, Aikawa E, et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for

Extracellular Vesicles and update of the MISEV2014 guidelines. *J Extracell Vesicles*. 2018;7(1). doi:10.1080/20013078.2018.1535750

22. Van Deun J, Mestdagh P, Agostinis P, et al. EV-TRACK: Transparent reporting and centralizing knowledge in extracellular vesicle research. *Nat Methods*. 2017;14(3):228-232. doi:10.1038/nmeth.4185

23. Wiśniewski JR, Zougman A, Mann M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J Proteome Res*. 2009;8(12):5674-5678. doi:10.1021/pr900748n

24. Cuklina J, Lee CH, Williams EG, et al. Systematic overview of batch effects in proteomics. In: *Computational Challenges in Biomarker Discovery from High-Throughput Proteomic Data*. ; 2018. doi:https://doi.org/10.3929/ethz-b-000307772

25. Pontén F, Jirström K, Uhlen M. The Human Protein Atlas - A tool for pathology. *J Pathol*. 2008;216(4):387-393. doi:10.1002/path.2440

26. Uhlén M, Fagerberg L, Hallström BM, et al. Tissue-based map of the human proteome. *Science (80-)*. 2015;347(6220). doi:10.1126/science.1260419

27. Keen JC, Moore HM. The genotype-tissue expression (GTEx) project: Linking clinical data with molecular analysis to advance personalized medicine. *J Pers Med*. 2015;5(1):22-29. doi:10.3390/jpm5010022

28. Yu NYL, Hallström BM, Fagerberg L, et al. Complementing tissue characterization by integrating transcriptome profiling from the human protein atlas and from the FANTOM5 consortium. *Nucleic Acids Res*. 2015;43(14):6787-6798. doi:10.1093/nar/gkv608

29. MacParland SA, Liu JC, Ma XZ, et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat Commun*. 2018;9(1). doi:10.1038/s41467-018-06318-7

30. Ma L, Hernandez MO, Zhao Y, et al. Tumor Cell Biodiversity Drives Microenvironmental Reprogramming in Liver Cancer. *Cancer Cell*. 2019;36(4):418-430.e6. doi:10.1016/j.ccell.2019.08.007

31. Ma L, Wang L, Khatib SA, et al. Single-cell atlas of tumor cell evolution in response to therapy in hepatocellular carcinoma and intrahepatic cholangiocarcinoma. *J Hepatol*. Published online 2021. doi:10.1016/J.JHEP.2021.06.028

32. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using “ggplot2.” Published online 2020. https://cran.r-project.org/package=survminer

33. Boberg KM, Bergquist A, Mitchell S, et al. Cholangiocarcinoma in primary sclerosing cholangitis: Risk factors and clinical presentation. *Scand J Gastroenterol*. 2002;37(10):1205-1211. doi:10.1080/003655202760373434

34. Chapman MH, Webster GJM, Bannoo S, Johnson GJ, Wittmann J, Pereira SP. Cholangiocarcinoma and dominant strictures in patients with primary sclerosing cholangitis: A 25-year single-centre experience. *Eur J Gastroenterol Hepatol*. 2012;24(9):1051-1058. doi:10.1097/MEG.0b013e3283554bbf

35. Bergquist A, Ekbom A, Olsson R, et al. Hepatic and extrahepatic malignancies in primary sclerosing cholangitis. *J Hepatol*. 2002;36(3):321-327. doi:10.1016/S0168-8278(01)00288-4

36. Boonstra K, Weersma RK, van Erpecum KJ, et al. Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. *Hepatology*.

2013;58(6):2045-2055. doi:10.1002/hep.26565

37. Vedeld HM, Folseraas T, Lind GE. Detecting cholangiocarcinoma in patients with primary sclerosing cholangitis – The promise of DNA methylation and molecular biomarkers. *JHEP Reports*. 2020;2(5):100143. doi:10.1016/j.jhepr.2020.100143

38. Johnstone RM, Adam M, Hammonds JR, Orro L, Turbide C. *THE JOURNAL OF BIOLOGICAL CHEMISTRY Vesicle Formation during Reticulocyte Maturation ASSOCIATION OF PLASMA MEMBRANE ACTIVITIES WITH RELEASED VESICLES (EXOSOMES)**. Vol 262.; 1987. doi:10.1016/S0021-9258(18)48095-7

39. Zhang Y, Zhang J, Chen X, Yang Z. Polymeric immunoglobulin receptor (Pigr) exerts oncogenic functions via activating ribosome pathway in hepatocellular carcinoma. *Int J Med Sci*. 2021;18(2):364-371. doi:10.7150/ijms.49790

40. Ai J, Tang Q, Wu Y, et al. The role of polymeric immunoglobulin receptor in inflammation-induced tumor metastasis of human hepatocellular carcinoma. *J Natl Cancer Inst*. 2011;103(22):1696-1712. doi:10.1093/jnci/djr360

41. Zhang X, Wang F, Huang Y, et al. FGG promotes migration and invasion in hepatocellular carcinoma cells through activating epithelial to mesenchymal transition. *Cancer Manag Res*. 2019;11:1653-1665. doi:10.2147/CMAR.S188248

42. Chang YH, Lee SH, Liao IC, Huang SH, Cheng HC, Liao PC. Secretomic analysis identifies alpha-1 antitrypsin (A1AT) as a required protein in cancer cell migration, invasion, and pericellular fibronectin assembly for facilitating lung colonization of lung adenocarcinoma cells. *Mol Cell Proteomics*. 2012;11(11):1320-1339. doi:10.1074/mcp.M112.017384

43. Zhang Y, Qiao HX, Zhou YT, Hong L, Chen JH. Fibrinogen-like-protein 1 promotes the invasion and metastasis of gastric cancer and is associated with poor prognosis. *Mol Med Rep*. 2018;18(2):1465-1472. doi:10.3892/mmr.2018.9097

44. Wang J, Wei W, Tang Q, et al. Oxysophocarpine suppresses hepatocellular carcinoma growth and sensitizes the therapeutic blockade of anti-Lag-3 via reducing FGL1 expression. *Cancer Med*. 2020;9(19):7125-7136. doi:10.1002/cam4.3151

45. Han Y, Wang J, Xu B. Novel biomarkers and prediction model for the pathological complete response to neoadjuvant treatment of triple-negative breast cancer. *J Cancer*. 2021;12(3):936-945. doi:10.7150/JCA.52439

46. Ni W, Song E, Gong M, Li Y, Yao J, An R. Downregulation of lncRNA SDPR-AS is associated with poor prognosis in renal cell carcinoma. *Onco Targets Ther*. 2017;10:3039-3047. doi:10.2147/OTT.S137641

47. Wang QX, Chen ED, Cai YF, et al. Serum deprivation response functions as a tumor suppressor gene in papillary thyroid cancer. *Clin Genet*. 2019;96(5):418-428. doi:10.1111/cge.13609

48. Luo X, Peng S, Ding S, et al. Prognostic values, ceRNA network, and immune regulation function of SDPR in KRAS-mutant lung cancer. *Cancer Cell Int*. 2021;21(1). doi:10.1186/s12935-021-01756-8

49. Jing W, Luo P, Zhu M, Ai Q, Chai H, Tu J. Prognostic and Diagnostic Significance of SDPR-Cavin-2 in Hepatocellular Carcinoma. *Cell Physiol Biochem*. 2016;39(3):950-960. doi:10.1159/000447803

PART V

GENERAL DISCUSSION

Cholangiocarcinoma (CCA) represents a substantial clinical, social, and economic global problem. Although considered a rare disease, its incidence and mortality rates are increasing worldwide, pinpointing the necessity to study in detail this cancer as a way to improve patients' welfare and outcome. The accuracy of current imaging techniques and circulating unspecific tumor markers for CCA diagnosis is far from being satisfactory, resulting frequently in unconcreted or late-stage diagnosis, when hardly any curative treatments may be applicable and palliative treatment turns out to be the last resource to slightly extend patients' survival.

CCA management is currently associated with a considerable economic burden that has risen in the past years. A retrospective analysis of the direct medical costs of 23,315 patients with iCCA between 2000 and 2018 in Spain revealed that the mean annual direct medical cost of secondary care was €9,417 per patient, with estimating a boost in total costs if incidence rates follow the same increasing trends observed in the last decades.[1] Additionally, a retrospective analysis in the United States including 1,298 patients with CCA who experienced failure of a first-line therapy containing either gemcitabine or fluorouracil (*i.e.*, patients with previously-treated advanced CCA) disclosed that the total mean of CCA-related healthcare inflation-adjusted costs per patient per month were $7,743.[2] Unfortunately, all these costs are mostly used to provide a palliative treatment of the best supportive care for these patients, which barely increases their outcome. Therefore, improvement in diagnostic techniques that allow the early detection of CCA would substantially reduce the economic cost of this cancer, maximizing the efficacy of the current therapeutic regimens.

The unmet clinical need of improved CCA detection may be solved by using novel molecular non-invasive biomarkers, such as circulating RNA and proteins. The discovery of safe and accurate tools for the non-invasive early tumor identification or prediction of development in patients at risk will lead to a greater number of patients suitable for potentially curative surgical treatments, improving the outcomes and minimizing the burden of illness associated with CCA. Particularly, the need of accurate early CCA biomarkers is extremely crucial for patients at high risk of developing CCA, where the current imaging techniques are unable to demonstrate the neoplastic nature of the results, such as in patients with PSC. The clinical application of new predictive and/or diagnostic precise strategies would allow establishing screening strategies to monitor these patients, minimizing CCA-related clinical, social and economic issues worldwide.

Liquid biopsy, the concept of evaluating health and/or diseased solid tissues by sampling circulating biofluids, is becoming a very popular approach that aims to replace the invasive methods used in current diagnostic protocols. Liquid biopsy offers the possibility to access the state of health of specific tissues in the body by detecting their released molecules in easily accessible body fluids, representing a simple and practical alternative to invasive biopsies, presumably evolving into the future “gold standard” technique for disease diagnosis.[3] On account of being minimally invasive and not requiring collecting samples from the solid tissues, the sampling procedures are well tolerated and allow obtaining longitudinal samples or repeated measurements. Therefore, the information of secreted materials is potentially extremely valuable and the barrier to obtain those materials is very low. Notwithstanding, currently, the great challenge for liquid biopsy is that the abundance of molecules released into biofluids by diseased tissues is relatively low, so identifying and quantifying them requires a great technological effort.[4] Another matter is that liquid biopsy biomarkers, which are found in relatively low amounts in biofluids, are supposed to allude the particular organ/tissue where the disease process is ongoing. For this matter, the Human Protein Atlas (HPA) represents a useful repository of information about mRNA and protein expression across not only multiple healthy tissues but also the most common cancers.[5] HPA contains mRNA sequencing data of specific tissues providing comprehensive views of gene expression, indicating which proteins are expected to be found in the different organs of the body. However, organs and tissues in healthy and diseased states typically represent complex diverse masses composed by distinct interacting cells which may differ widely in ontogeny, function and gene expression. On this matter, the recently developed single-cell sequencing (scRNA-seq) technology is a valuable tool for dissecting cellular heterogeneity in complex systems and provides information about individual cells that comprise tissues and organs.[6] Following scRNA-seq methodology, with the aim of mapping every cell type in the human body, an open global international collaborative consortium called the Human Cell Atlas (HCA) was launched in 2016.[7] Recently, the HCA consortium has published the first approach to the initiative by integrating a wide range of human tissue single-cell transcriptomic datasets and constructing a human cell landscape at single-cell level.[8] This approach would lead to major advances in the way illnesses are diagnosed and treated.

With regard to biological fluids as a source of liquid biopsies, serum and urine are the most widely handled biofluids owing to the fact they are quite easily removed by low-invasive or non-invasive procedures along with the fact that they provide enough volume for the required analysis. Both serum and urine fluids have their advantages and

limitations for liquid biopsy and may provide complementary information. Sampling of blood is minimally invasive and it requires healthcare professional experts while the acquisition of urine is simple and non-invasive, potentially being conducted by the patient itself. Besides, blood biomolecules can provide information at systemic level and, as a result of tumor irrigation, the molecular profiling of blood presumably mimic tumor-secreted molecules. However, urinary excretion is physiologically influenced by the glomerular filtration, tubular reabsorption and tubular secretion, potentially inducing variations in the abundance and composition of molecules secreted by tumor cells.[9]

Serum and urine biofluids contain a wide range of components including circulating cells, cell-free DNAs, mRNAs, miRNAs, proteins, metabolites and EVs.[10–13] Among the molecular composition of biofluids, EVs are emerging as worthwhile biomarker-packaging liquid biopsy tools for the diagnosis of several diseases. In this sense, we hypothesized that circulating EVs might be a valuable source of novel accurate biomarkers that would allow the prediction, early diagnosis and prognosis estimation of CCA. In the studies included in this dissertation, we have described the presence of RNA molecules and proteins in serum and/or urine of ultracentrifugation-isolated EVs by microarray profiling and by mass-spectrometry high-throughput proteomics, respectively. The abundance of some EV molecules differed in patients with CCA (with or without PSC etiology) compared to healthy individuals and patients with UC or PSC, pinpointing some biomolecules as promising biomarker candidates for CCA prediction, early detection and survival estimation. The fact that numerous circulating biomarkers were also expressed in CCA tumor tissues and specifically in malignant cholangiocytes within the tumor pointed out the neoplasia-associated essence of candidate circulating biomarkers, which constitutes a fundamental characteristic for liquid biopsy.

The followed strategy of analysing serum and urine EVs by high-throughput omics rather than unprocessed whole serum or urine omic profiling was conducted for several reasons. Firstly, a fundamental consideration for protein- and/or nucleotide-based liquid biopsies is that really profitable cancer biomarkers may turn out to be present at very low levels in circulating biofluids. Biomolecules that are found at high concentrations in circulation from normal individuals are unlikely to be noticeably increased in patients with a limited tumor mass, a characteristic of early tumor disease stages.[3] In the current omics era, analysis of serum biofluids for proteomic-based biomarker discovery is highly challenging due to the complexity and wide dynamic range of their proteomes, where the concentrations of the least abundant and the most abundant proteins can differ by as much as ten orders of magnitude.[14] For instance, albumin is the most abundant protein in serum, accounting for around 55% of the total protein content, and together with IgG,

IgA, haptoglobin, transferrin and alpha-1-antitrypsin, they constitute approximately 85% of the total proteome. Notably, in high-throughput proteomics, highly abundant proteins tend to obscure the detection of potential biomarkers that are usually at lower concentrations. Thereby, the vesicular nature of EVs is one of the central methods used to separate the highly abundant proteins and to enhance the identification and detection capacity for the least abundant proteins in serum. Therefore, vesicle enrichment is profitable for mass spectrometry-based analysis, as EV-subproteome represents only a very small component of all serum proteins.[4]

Secondly, a biomolecule may be considered a good candidate biomarker for liquid biopsy if it could be secreted to the extracellular media and enter the circulation. It is known that EVs originate in multivesicular bodies (*i.e.*, exosomes) or thorough plasma membrane budding (*i.e*., microvesicles) of the cells and they are secreted into the extracellular matrix, reaching diverse biofluids.[10,15] Accordingly, EVs can provide an accurate representation of the condition of the cell they are released from, likely containing key information concerning the state of the cell of origin. Conveniently, due to their vesicular nature which confers them with unique physicochemical properties, EVs can be efficiently separated from the most abundant soluble serum proteins or nucleic acids by many isolation approaches, designating them as available packages of unique information.[16]

Last but not least, owing to the fact that RNAs are relatively unstable molecules susceptible to degradation by ribonucleases found in circulation, EV encapsulation of ribonucleic acids prevents them from nuclease fragmentation or degradation activity, increasing their biofluid existence period.[17,18]

Regarding EVs and high-throughput omic technologies, in the last years, scientific publications discovering non-invasive disease biomarkers in biofluid EVs are undergoing an exponential increase while the clinical application of discovered EV biomarkers are still far from being realistic.[10,12,13,19] The rationability of this huge gap between the number of omics-based biomarkers found in basic research literature and those introduced into the clinics may be related to the arduousness of EV isolation techniques, the time consuming procedures, the specific equipment requirements, the wide range of EV isolation protocols which influence omic results, the lack of consensus strategies in the field, the high inter-procedure and intra-procedure variability of isolation methodologies, the needed sample volume, as well as the low technical reproducibility, among others. Therefore, the consideration of using proteomic and transcriptomic high-throughput platforms for the molecular profiling of isolated EVs to identify candidate early non-

invasive biomarkers and subsequent validation of biomarkers in unprocessed samples with simple straightforward techniques represents a rational combination approach for biomarker development and clinical application. In this regard, quantitative polymerase chain reaction (qPCR) and immunoaffinity-based protein assays have displayed as the preferred analytical platforms by the scientific community for the validation of ribonucleotide and protein biomarkers in unprocessed biofluids, thus, avoiding time-consuming steps and equipment demands of EV isolation procedures.

For nucleotide quantification, high sensitivity and precision is typically associated with qPCR technology. The chances of a successful clinical translation of a RNA biomarker will depend on the reliable measurements performed by this robust assay. In line with this technique, droplet digital polymerase chain reaction (ddPCR) has recently been developed. ddPCR is an ultrasensitive detection technique based on sample partitioning into thousands of nanoliter-sized droplets where individual PCR reactions take place, allowing for the detection of very low abundance molecular targets with extremely high sensitivity.[20,21] The main advantages of ddPCR *versus* the gold standard RT-qPCR method include a superior sensitivity and accuracy for ddPCR, as well as the ability to perform an absolute quantification without standard curves. Due to its reliability, accuracy and allowance of multiplexing strategies which reduces the required sample amount, ddPCR is usually applied as a validation technique.[22,23]

Affinity-based protein assays, mostly using antibodies as target-binding reagents, offer the greatest detection sensitivity for protein validation assays. They are relatively inexpensive with fast turn-around times. Current assays for sensitive protein detection in solution-phase rely on the classical sandwich enzyme-linked immunosorbent assay (ELISA), which requires a pair of antibodies to recognize the target protein, enhancing target recognition specificity.[3] The sensitivity of these affinity-based protein assays will be limited to the detection efficiency of recognition antibodies and the total amount of the biomarkers in unprocessed biofluids. Similar to ELISA assays, in the last years, proximity ligation or proximity extension assays (PEA) have also been introduced as protein validation strategies.[24] In PEA assays, target proteins are also recognized by two antibodies, keeping a similar specificity to sandwich ELISA assays, but those antibody pairs are conjugated to DNA oligonucleotides which following antibody proximity undergo DNA ligation and/or polymerization creating an amplifiable reporter DNA strand. The amplification of this DNA strand by RT-qPCR will exponentially enhance the sensitivity of this assay. PEA assays which convert protein identities into DNA sequence information *via* antibody-DNA conjugates can also be performed in multiplexing, detecting many different proteins in parallel and diminishing the required sample amount.

Despite technical validation of low abundant biomarkers in unprocessed biofluids by reproducible, sensitive and specific assays that are adaptable to routine clinical practice and have a timely turnaround, other considerations in the validation procedures would help the translation of the findings from bench to bedside. In particular, a high number of large prospective case-control studies involving a clearly defined set of patients would be essential, with matching confounding factors and including information of the clinical characteristics as completed as possible. The proper patients' selection and the inclusion of clinically relevant control groups will determine the results of the validation.

After biomarker discovery and analytical validation, the potential clinical utility and use should be evaluated in the pipeline of biomarker development.[25] Prospective clinical trials will evaluate the clinical usefulness of the biomarkers and after that, the biomarker test would have to get the regulatory approval, be commercialized and incorporated into clinical practice guidelines.

In the near future, healthcare challenges in terms of liquid biopsy such as increasing diagnostic accuracy and delivering individual care, rather than by individual biomarker testing, they will be uncovered by machine learning (ML) algorithms.[26] Single biomarkers in liquid biopsy often cannot accurately predict the state of a disease due to the variability in biomarker expression across both diseased and healthy individuals as well as due to the phenotypical heterogeneity of many diseases, including CCA.[27] To address this challenge, ML algorithms outperforms the sensitivity and specificity of individual biomarkers buy building models from sample multiple inputs or different biomarker measurements, thus making predictions and improving the diagnostic performance. ML encompasses a set of computational techniques widely applied in many fields to reduce large numbers of measurements into lower-dimensional outputs that are more useful. ML-based softwares could be used in screening programs to identify and monitor high-risk patients and they could also decrease the current cost of diagnostic strategies for CCA. In order to turn data into meaningful insights, different ML-based models should be done, in parallel with a proper selection of the individuals. The efficacy of a ML algorithm is mostly dependent on the quality of the training dataset. Quality, robustness and reproducibility of the underlying data will determine the output result and discrepancies in the data collection process, imperfections of standardization and incorrect labelled cases will inevitably limit the ML-based software's accuracy. Future integration of ML algorithms into healthcare should go along with physicians' intelligence for a ML-assisted but not ML-driven clinical practice.[26]

In summary, in this study, biomarkers and biomarker-combining ML models of ribonucleotides or proteins have been proposed in serum and urine EVs that allowed the prediction, early detection and progression of CCAs in general and in CCAs arising in a PSC background. The consideration of using biofluid-isolated EV proteomic and transcriptomic platforms provided an important avenue for the discovery of early disease biomarkers. The validation of these early CCA biomarkers along with others that would arise in the future due to ongoing studies as well as logistic models combining multi-omic data would allow malignant processes to be interrupted before metastatic spread, resulting in more efficient treatments for the individuals concerned, and subsequently impacting in healthcare cost minimization and social welfare. Furthermore, successfully developed assays would help for effective population screening at early signs of malignization to enable curative surgery and anticipate the appearance of late symptoms. This global benefit would only be possible by the development of international collaborative consortia such as the European Network for the Study of Cholangiocarcionoma (ENS-CCA), the establishment of collaborative and accessible open databases that shares information to the scientific community (*e.g.*, MarkerDB, the online database of molecular biomarkers), the use of international patient registries (*e.g.*, the ENS-CCA) as well as by complex logistic strategies for sample collecting, stocking and distributing in national and international biobanks. Larger sample collections would be required for biomarker validation, with high number of control individuals whose biomarker levels would contribute to define the normal concentration ranges of those molecules as well as samples collected before any onset of symptoms or at slight suspicion of the disease. In this sense, active awareness of patients to get involved into research by donating samples is crucial and foundations such as the US Cholangiocarcionma Foundation and the AMMF - The UK Cholangiocarcinoma Charity play a fundamental role in this aim. Finally, the success of these efforts for accurate non-invasive early diagnosis of CCA will certainly make progress thorough the collaborative cross-disciplinary cooperation among, clinicians, basic and translational researchers, computational biologists, biostatistics and epidemiologists. Such partnerships would ultimately accelerate the translation of cutting-edge scientific discoveries from bench to bedside, thus, leading to improved patient care and outcome.

REFERENCES

1. Darbà J, Marsà A. Analysis of hospital incidence and direct medical costs of intrahepatic cholangiocarcinoma in Spain (2000-2018). *Expert Rev Pharmacoecon Outcomes Res*. 2021;21(3):425-431. doi:10.1080/14737167.2021.1842201

2. Chamberlain CX, Faust E, Goldschmidt D, et al. Burden of illness for patients with cholangiocarcinoma in the United States: a retrospective claims analysis. *J Gastrointest Oncol*. 2021;12(2):658-668. doi:10.21037/JGO-20-544

3. Landegren U, Hammond M. Cancer diagnostics based on plasma protein biomarkers: hard times but great expectations. *Mol Oncol*. 2021;15(6):1715-1726. doi:10.1002/1878-0261.12809

4. Burton JB, Carruthers NJ, Stemmer PM. Enriching extracellular vesicles for mass spectrometry. *Mass Spectrom Rev*. Published online 2021. doi:10.1002/MAS.21738

5. Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347(6220). doi:10.1126/SCIENCE.1260419

6. Tanay A, Regev A. Scaling single-cell genomics from phenomenology to mechanism. *Nature*. 2017;541(7637):331-338. doi:10.1038/NATURE21350

7. Regev A, Teichmann SA, Lander ES, et al. The Human Cell Atlas. *Elife*. 2017;6. doi:10.7554/ELIFE.27041

8. Han X, Zhou Z, Fei L, et al. Construction of a human cell landscape at single-cell level. *Nature*. 2020;581(7808):303-309. doi:10.1038/S41586-020-2157-4

9. Trovato FM, Tognarelli JM, Crossey MME, Catalano D, Taylor-Robinson SD, Trovato GM. Challenges of liver cancer: Future emerging tools in imaging and urinary biomarkers. *World J Hepatol*. 2015;7(26):2664-2675. doi:10.4254/WJH.V7.I26.2664

10. Lapitz A, Arbelaiz A, Olaizola P, et al. Extracellular Vesicles in Hepatobiliary Malignancies. *Front Immunol*. 2018;9(OCT). doi:10.3389/FIMMU.2018.02270

11. Olaizola P, Lee-Law PY, Arbelaiz A, et al. MicroRNAs and extracellular vesicles in cholangiopathies. *Biochim Biophys acta Mol basis Dis*. 2018;1864(4 Pt B):1293-1307. doi:10.1016/J.BBADIS.2017.06.026

12. Macias RIR, Kornek M, Rodrigues PM, et al. Diagnostic and prognostic biomarkers in cholangiocarcinoma. *Liver Int*. 2019;39 Suppl 1(S1):108-122. doi:10.1111/LIV.14090

13. Macias RIR, Banales JM, Sangro B, et al. The search for novel diagnostic and prognostic biomarkers in cholangiocarcinoma. *Biochim Biophys acta Mol basis Dis*. 2018;1864(4 Pt B):1468-1477. doi:10.1016/J.BBADIS.2017.08.002

14. Lee PY, Osman J, Low TY, Jamal R. Plasma/serum proteomics: depletion strategies for reducing high-abundance proteins for biomarker discovery. *Bioanalysis*. 2019;11(19):1799-1812. doi:10.4155/BIO-2019-0145

15. Van Niel G, D'Angelo G, Raposo G. Shedding light on the cell biology of extracellular vesicles. *Nat Rev Mol Cell Biol*. 2018;19(4):213-228. doi:10.1038/NRM.2017.125

16. Théry C, Witwer KW, Aikawa E, et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J Extracell vesicles*. 2018;7(1). doi:10.1080/20013078.2018.1535750

17. Pös O, Biró O, Szemes T, Nagy B. Circulating cell-free nucleic acids: characteristics and applications. *Eur J Hum Genet*. 2018;26(7):937-945. doi:10.1038/S41431-018-0132-4

18. Tzimagiorgis G, Michailidou EZ, Kritis A, Markopoulos AK, Kouidou S. Recovering circulating extracellular or cell-free RNA from bodily fluids. *Cancer Epidemiol*. 2011;35(6):580-589. doi:10.1016/J.CANEP.2011.02.016

19. Rodrigues PM, Vogel A, Arrese M, Balderramo DC, Valle JW, Banales JM. Next-Generation Biomarkers for Cholangiocarcinoma. *Cancers (Basel)*. 2021;13(13). doi:10.3390/CANCERS13133222

20. Pinheiro LB, Coleman VA, Hindson CM, et al. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal Chem*. 2012;84(2):1003-1011. doi:10.1021/AC202578X

21. Olmedillas-López S, Olivera-Salazar R, García-Arranz M, García-Olmo D. Current and Emerging Applications of Droplet Digital PCR in Oncology: An Updated Review. *Mol Diagn Ther*. Published online November 13, 2021. doi:10.1007/S40291-021-00562-2

22. Fitz NF, Wang J, Kamboh MI, Koldamova R, Lefterov I. Small nucleolar RNAs in plasma extracellular vesicles and their discriminatory power as diagnostic biomarkers of Alzheimer's disease. *Neurobiol Dis*. 2021;159. doi:10.1016/J.NBD.2021.105481

23. Campomenosi P, Gini E, Noonan DM, et al. A comparison between quantitative PCR and droplet digital PCR technologies for circulating microRNA quantification in human lung cancer. *BMC Biotechnol*. 2016;16(1). doi:10.1186/S12896-016-0292-7

24. Fredriksson S, Gullberg M, Jarvius J, et al. Protein detection using proximity-dependent DNA ligation assays. *Nat Biotechnol*. 2002;20(5):473-477. doi:10.1038/NBT0502-473

25. Quezada H, Guzmán-Ortiz AL, Díaz-Sánchez H, Valle-Rios R, Aguirre-Hernández J. Omics-based biomarkers: current status and potential use in the clinic. *Bol Med Hosp Infant Mex*. 2017;74(3):219-226. doi:10.1016/J.BMHIMX.2017.03.003

26. Christou CD, Tsoulfas G. Challenges and opportunities in the application of artificial intelligence in gastroenterology and hepatology. *World J Gastroenterol*. 2021;27(37):6191-6223. doi:10.3748/WJG.V27.I37.6191

27. Ko J, Baldassano SN, Loh PL, Kording K, Litt B, Issadore D. Machine learning to detect signatures of disease in liquid biopsies - a user's guide. *Lab Chip*. 2018;18(3):395-405. doi:10.1039/C7LC00955K

PART V

CONCLUSIONS

1. Serum and urine EVs isolated by serial differential ultracentrifugation steps are mainly enriched in exosomes and microvesicles, characterized by the typical round morphology, a size around 200nm and an enrichment in EV markers CD63 and CD81.

2. The transcriptomic and proteomic biofluid EV profiling revealed that among all the identified RNA and protein molecules, some specific molecule levels were differentially detected in patients with CCA compared to healthy individuals, patients with UC and patients with isolated PSC. The proteomic profile also differed in some proteins when comparing patients with CCA with or without a PSC background, as well as between patients with PSC who develop CCA in the follow-up and patients with non-malignant PSC.

3. Biomarker performance evaluating metrics revealed that biofluid EVs contain accurate RNA and/or protein biomarkers for bile duct cancers, with a higher sensitivity and specificity than the serum tumor marker CA19-9. EV-related biomarkers revealed the capacity for the prediction of CCA development in patients with PSC, for the early tumor diagnosis (in patients with PSC background, in patients without PSC etiology and in CCAs regardless disease etiology) and for the prognosis estimation of patients with CCA.

4. Certain candidate EV-related biomarkers are expressed in CCA tumors, preferentially in malignant cholangiocytes. This fact suggests the possibility that the newly described biofluid EV biomarkers are produced by CCA tumors and released to serum and/or urine, mirroring tumor behaviour and emphasizing this novel liquid biopsy strategy. Moreover, these results pave the path not only for the discovery of new biomarkers but also for new potential therapeutic targets, since they may participate in disease pathogenesis.

5. Machine learning algorithms rather than individual biomarkers seem to be more powerful in terms of diagnostic and prognostic testing parameters, which include sensitivity, specificity, positive predictive value, negative predictive value, accuracy, odds ratio and hazard ratio. These machine learning algorithms allow the correct prediction, diagnosis or prognosis in a wider population, diminishing the false positive or negative results that may arise from individual biomarker testing.

6. A logistic model considering serum EV mRNA levels of *CMIP*, *NME1* and *CKS1B* discriminated patients with CCA from patients with PSC, UC and healthy individuals with 100% of accuracy.

7. For patients with PSC, an algorithm considering serum EV protein levels of HV308, KAIN, HEMO, A2MG, VWF, APOA1, TLN1 and IMA8 allowed the diagnosis of CCA with 100% of accuracy, also at early tumor stages. In addition, a logistic model combining only PLCH1 and FGL1 diagnosed patients with early-stage PSC-CCA compared to individuals with PSC with 91% of accuracy.

8. For patients without PSC, a logit model combining the protein biomarkers SAMP, ACTC, RELN, APOA1 and KLKB1 enabled 100% of sensitivity and specificity in the diagnosis of CCA compared to healthy individuals. Furthermore, an algorithm combining only SAMP and A1AT revealed a sensitivity of 77% and specificity of 90% when discriminating patients with CCA compared to healthy individuals.

9. A panel considering serum EV levels of FCN2, SDPR and FA9 predicted overall survival of patients with CCA. High levels of these 3 proteins was associated with a median overall survival of more than 10 years, while the low abundance in 2 or 3 biomarkers leads to a 30-fold death risk and a median survival of less than 7 months.

PART VI

APPENDIX

ABBREVIATIONS

1433Z	14-3-3 protein zeta/delta
4F2hc	4F2 cell-surface antigen heavy chain
^{90}Y-TARE	Yttrium90 transarterial radioembolization
A1AG1	Alpha-1-acid glycoprotein 1
A1AT	Alpha-1-antitrypsin
A2M	Alpha-2-macroglobulin gene
A2MG	Alpha-2-macroglobulin protein
ACTB	Actin cytoplasmic 1
ACTC	Actin alpha cardiac muscle 1
ACTN1	Alpha-actinin-1
ADA10	Disintegrin and metalloproteinase domain-containing protein 10
ADAM10	Metalloproteinase domain-containing protein 10 gene
ADIPO	Adiponectin
ADMSCs	Adipose tissue-derived mesenchymal stem cells
AFP	Alpha fetoprotein
AI	Accuracy index
AJCC	American Joint Committee on Cancer
ALB	Albumin gene
ALBU	Albumin protein
ALP	Alkaline phosphatase
ALPL	Alkaline phosphatase, biomineralization associated gene
ALT	Alanine transaminase
AMBP	Protein AMBP
AMPN	Aminopeptidase N
ANOVA	Analysis of variance
APCs	Antigen presenting cells
APCS	Serum amyloid P-component
APOA1	Apolipoprotein A-I
APOA2	Apolipoprotein A-II
APOC1	Apolipoprotein C-I
APOE	Apolipoprotein E
APOF	Apolipoprotein F
APOL1	Apolipoprotein L1
AST	Aspartate transaminase
ATF4	Activating transcription factor 4
ATP5EP2	ATP synthase F1 subunit epsilon pseudogene 2
AUC	Area under the receiver operating characteristic curve
BWS	Beckwith-Widemann syndrome
C1S	Complement C1s subcomponent
C9	Complement C9
CA19-9	Carbohydrate antigen 19-9
CAF	Cancer-associated fibroblast
CAGE	Cap analysis of gene expression
CBD	Common bile duct
CCA	Cholangiocarcinoma
CCNG1	Cyclin G1
CD163	Scavenger receptor cysteine-rich type 1 protein M130
CD19	B-lymphocyte antigen CD19
CD34	Hematopoietic progenitor cell antigen CD34
CD3D	T-cell surface glycoprotein CD3 delta chain

CD3E	T-cell surface glycoprotein CD3 epsilon chain
CD3G	T-cell surface glycoprotein CD3 gamma chain
CD68	Macrosialin
CD79A	B-cell antigen receptor complex-associated protein alpha chain
CDC	Cell division control
CDK	Cyclin-dependent kinases
CDS1	CDP-diacylglycerol synthase 1
CEA	Carcinoembryonic antigen
CEUS	Contrast-enhanced ultrasonography
CFL1	Cofilin 1 gene
CHB	Chronic hepatitis B
CI	Confidence interval
CKS1B	Cyclin-dependent kinases regulatory subunit 1
CLIP3	CAP-Gly domain containing linker protein 3
CME	Clathrin-mediated endocytosis
CMIP	C-Maf inducing protein
CO4A	Complement C4-A
CO4B	Complement C4-B
CO9	Complement component C9 protein
COF1	Cofilin-1
COL1A2	Collagen Type I Alpha 2 Chain
CT	Computed tomography
dCCA	Distal cholangiocarcinoma
DCN	Decorin
DCs	Dendritic cells
ddPCR	Droplet digital polymerase chain reaction
DFMO	Difluoromethylornithine
DLS	Dynamic light scattering
DMA	Dimethyl amiloride
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EASL	European Association for the Study of the Liver
eCCA	Extrahepatic cholangiocarcinoma
ECOG-PS	Eastern Cooperative Oncology Group Performance Status
EGFR	Epidermal growth factor receptor
EIPA	5-ethyl-N-isopropyl amiloride
ELISA	Enzyme-linked immunosorbent assay
EMT	Epithelial-mesenchymal transition
ENS-CCA	European Network for the Study of Cholangiocarcinoma
EPCAM	Epithelial cell adhesion molecule
ER	Endoplasmic reticulum
ERCP	Endoscopic retrograde cholangiopancreatography
ESCRT	Endosomal sorting complex required for transport
EUS	Endoscopic ultrasonography
EVs	Extracellular vesicles
FA9	Coagulation factor IX
FAM107B	Family with sequence similarity 107 member B
FANTOM5	Functional Annotation of Mammalian Genomes 5
FAP	Familial adenomatous polyposis
FASP	Filter aided sample preparation

FBS	Fetal bovine serum
FCGBP	IgGFc-binding protein
FCN1	Ficolin-1
FCN2	Ficolin-2
FCRL5	Fc receptor-like protein 5
FGB	Fibrinogen beta chain
FGFR2	Fibroblast growth factor receptor 2
FGG	Fibrinogen gamma chain
FGL1	Fibrinogen-like protein 1
FIBA	Fibrinogen alpha chain
FIBB	Fibrinogen beta chain
FIBG	Fibrinogen gamma chain
FISH	Fluorescence in situ hybridization
FLNA	Filamin-A
FOLFOX	Folinic acid, 5-fluorouracil and oxaliplatin
FRIL	Ferritin light chain protein
FunRich	Functional Enrichment analysis tool
FUT3	Fucosyltransferase 3 (Lewis Blood Group)
FXYD2	FXYD domain containing ion transport regulator 2
GAD1	Glutamate decarboxylase 1
GC	Vitamin D-binding protein coding gene
GELS	Gelsolin
GGT	Gamma-glutamyltransferase
GO	Gene ontology
GPX3	Glutathione peroxidase 3
GRP78	78 kDa glucose-regulated protein, endoplasmic reticulum chaperone BiP
GSTM1	Glutathione S-Transferase Mu 1
GSTT1	Glutathione S-Transferase Theta 1
GTEx	Genotype-Tissue Expression
GWAS	Genome-wide association study
HB	Hepatoblastoma
HBV	Hepatitis B virus
HCA	Human Cell Atlas
HCC	Hepatocellular carcinoma
HCG4	HLA complex group 4
HCV	Hepatitis C virus
HEMO	Hemopexin
HGF	Hepatocyte growth factor
hOGG1	Human oxoguanine glycosylase 1
HPA	Human Protein Atlas
HPC	Hepatic progenitor cell
HPR	Haptoglobin-related protein
HPTR	Haptoglobin-related protein coding gene
HPX	Hemopexin gene
HR	Hazard ratio
HSP7C	Heat shock cognate 71 kDa protein
HSPs	Heat shock proteins
HSPG	Heparan sulfate proteoglycans
hTERT	Telomerase reverse transcriptase
HV305	Immunoglobulin heavy variable 3-13

HV308	Immunoglobulin heavy variable 3-30
IBD	Inflammatory bowel disease
ICAM-1	Intercellular adhesion molecule 1
iCCA	Intrahepatic cholangiocarcinoma
ICD	International Classification of Disease
IDH1	Inhibitor of isocitrate dehydrogenase 1
IGF1R	Insulin-like growth factor 1 receptor
IGHA1	Immunoglobulin heavy constant alpha 1
IGHA2	Immunoglobulin heavy constant alpha 2
IGHG4	Immunoglobulin heavy constant gamma 4
IGLL5	Immunoglobulin lambda-like polypeptide 5
IL	Interleukin
ILVs	Intraluminal vesicles
IMA8	Importin subunit alpha-8
INO80D	INO80 complex subunit D
ITGB4	Integrin β4
ITIH1	Inter-alpha-trypsin inhibitor heavy chain H1
ITIH2	Inter-alpha-trypsin inhibitor heavy chain H2
ITIH4	Inter-alpha-trypsin inhibitor heavy chain H4
JAM2	Junctional adhesion molecule B
K1C9	Keratin, type I cytoskeletal 9
K22E	Keratin, type II cytoskeletal 2 epidermal protein
K2C6A	Keratin, type II cytoskeletal 6A
KAIN	Kallistatin
KLKB1	Plasma kallikrein
KLRF1	Killer cell lectin-like receptor subfamily F member 1
KRT19	Keratin, type I cytoskeletal 19
KRT2	Keratin 2
KRT7	Keratin, type II cytoskeletal 7
KV121	Immunoglobulin kappa variable 1D-33
KV306	Immunoglobulin kappa variable 3-15
LAC2	Immunoglobulin lambda constant 2
LBP	Lipopolysaccharide-binding protein
LC	Liver cancer
LC	Liver cirrhosis
LC	Liquid chromatography
LDH4	Lactate dehydrogenase 4
LFG3	Protein lifeguard 3
LG3BP	Galectin-3 binding protein
linc-ROR	Long intergenic non-coding RNA regulator of reprogramming
lncRNA	Long non-coding RNA
LRP1	Prolow-density lipoprotein receptor-related protein 1
LT	Liver transplantation
LUM	Lumican
LV301	Immunoglobulin lambda variable 3-1
LV403	Immunoglobulin lambda variable 4-3
MβCD	Methyl-β-cyclodextrin
MALAT1	Metastasis associated lung adenocarcinoma transcript 1

MAP6D1	MAP6 domain containing 1
MHC	Major histocompatibility complex
MICB	MHC class I polypeptide-related sequence B
miRNA	MicroRNA
miscRNA	Miscelaneous RNA
ML	Machine learning
MMPs	Matrix metalloproteinases
MMRN1	Multimerin-1
mOS	Median overall survival
MRCP	Magnetic resonance cholangiopancreatography
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
MRP2/ABCC2	Multidrug resistance-associated protein 2
MS	Mass spectrometry
MSCs	Mesenchymal stem cells
MSR1	Macrophage scavenger receptor 1
MT1F	Metallothionein 1F
MTHFR	5,10-methylenetetrahydrofolate reductase
MUTYH	MutY homolog
MVE	Multivesicular endosome
MVs	Microvesicles
MYL6B	Myosin light chain 6B
NAFLD	Non-alcoholic fatty liver disease
NASH	Non-alcoholic steatohepatitis
NCAM1	Neural cell adhesion molecule 1
ncRNA	Non-coding RNA
NBD	Normal bile ducts
NHCs	Normal human cholangiocytes
NK	Natural killer
NKG2D	Natural killer cell receptor G2D
NKT	Natural killer T cell
NLR	Negative likelihood ratio
NME1	Nucleoside diphosphate kinase 1
nSMase2	Neutral type II sphingomyelinase
NPV	Negative predictive value
NTA	Nanoparticle tracking analysis
OIT3	Oncoprotein-induced transcript 3
OR	Odds ratio
OR4F3	Olfactory receptor family 4 subfamily F member 3
OS	Overall survival
PAFs	Para-cancer fibroblasts
Pan-CCAs	CCAs regardless etiology
PBMCs	Peripheral blood mononuclear cells
PBS	Phosphate-buffered saline
PBX3	Pre-B-cell leukemia transcription factor 3
PCA	Principal component analysis
PCA3	Prostate cancer associated 3
pCCA	Perihilar cholangiocarcinoma
PD-1	Programmed cell death protein I
PDT	Photodynamic therapy

PEA	Proximity extension assays
PET	Positron emission tomography
PFA	Paraformaldehyde
PHGDH	Phosphoglycerate dehydrogenase
PI3K	Phosphoinositide-3-kinase
PIGR	Polymeric immunoglobulin receptor
PLCH1	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase eta-1
PLD2	Phospholipase D2
PLEK	Pleckstrin
PLR	Positive likelihood ratio
PON1	Paraoxonase 1
PMS2L4	PMS1 homolog 2 mismatch repair system component pseudogene 4
PPBT	Alkaline phosphatase, tissue-nonspecific isozyme
PPV	Positive predictive value
PS	Phosphatidylserine
PSA2	Proteasome subunit alpha type-2
PSA3	Proteasome subunit alpha type-3
PSB3	Proteasome subunit beta type-3
PSC	Primary Sclerosing Cholangitis
PSC-CCA	Concomitant primary sclerosing cholangitis and cholangiocarcinoma
PTC	Percutaneous transhepatic cholangiography
qPCR	Quantitative polymerase chain reaction
RAB10	Ras-related protein Rab-10
RELN	Reelin
RFA	Radiofrequency ablation
RFFL	Ring finger and FYVE like domain containing E3 ubiquitin protein ligase
RNA	Ribonucleic acid
RNU11	U11 small nuclear
RRAGD	Ras-related GTP binding D
SAMP	Serum amyloid P-component protein
SBRT	Stereotactic body radiotherapy
scRNA-seq	Single-cell RNA sequencing
SDPR	Caveolae-associated protein 2
SDS-PAGE	Sodium dodecyl sulfate-polyacrilamide gel electrophoresis
SEER	Surveillance, Epidemiology, and End Results
SEN	Sensitivity
SERPINA1	Serpin family A member 1
SERPINA10	Serpin family A member 10
SERPINA4	Kallistatin gene
SERPINB1	Serine protease inhibitor B1
SIRT	Selective internal radiation therapy
SL	Surrounding liver
snoRNA	Small nucleolar RNA
SNP	Single nucleotide polymorphisms
snRNA	Small nuclear RNA
SPE	Specificity
SVEP1	Sushi, von Willebrand factor type A, EGF and pentraxin domain containing protein **1**
TACE	Transarterial chemoembolization

TAMPs	Tumor-associated microparticles
TBS	Tris-buffered saline
TBS-T	Tris-buffered saline with 0.1% Tween 20
TCGA	The Cancer Genome Atlas
TCTEX1D2	Tctex1 domain containing 2
TEM	Transmission Electron Microscopy
TFR1	Transferrin receptor protein 1
TFRC	Transferrin receptor gene
ThO_2	Thorium dioxide
TLN1	Talin-1
TME	Tumor microenvironment
TNM	Tumor, node and metastasis tumor classification system
TRIM33	Tripartite motif containing 33
TSN9	Tetraspanin-9 protein
TTYH3	Protein tweety homolog 3
UA	Uranyl acetate
UBE2C	Ubiquitin conjugating enzyme E2 C
UC	Ulcerative colitis
US	Ultrasonography
VCAM1	Vascular cell adhesion molecule 1
VEGF	Vascular endothelial growth factor
VNN1	Pantetheinase
VTDB	Vitamin-D binding protein
vtRNA	Vault RNA
VTRNA1-1	Vault RNA 1-1
VWF	Von Willebrand factor protein/gene
WCE	Whole-cell extract
WHO	World Health Organization
YI	Youden index
ZNF266	Zinc finger protein 266

SUMMARY IN SPANISH (RESUMEN EN ESPAÑOL)

INTRODUCCIÓN

El colangiocarcinoma (CCA) agrupa un conjunto heterogéneo de tumores con características de diferenciación biliar que se originan a lo largo de la vía biliar, desde los canales de Hering hasta el conducto biliar común.[1–3] Los CCAs representan el 3% de los tumores gastrointestinales y el 10-15% de los casos de tumores hepáticos primarios, tratándose del segundo tipo de cáncer hepático primario más frecuente a nivel mundial, tras el carcinoma hepatocelular (CHC).[3]

El CCA es un cáncer poco frecuente, con una incidencia global menor a 6 casos por 100.000 habitantes, que afecta tanto a hombres como a mujeres en la década de los sesenta-setenta años, y cuya incidencia está en aumento.[3] La distribución geográfica mundial de este cáncer es muy heterogénea, con una incidencia muy elevada en regiones del sudeste asiático (ej. 85 casos por 100.000 habitantes en el nordeste de Tailandia) mientras que en algunos países occidentales es muy poco frecuente.[4] El crecimiento asintomático de los CCAs en estadios iniciales conlleva a un diagnóstico habitualmente tardío que, en combinación con su alta agresividad, las pocas opciones terapéuticas, su elevada quimio-resistencia y la alta tasa de recurrencia post-cirugía contribuyen a la alarmante tasa de mortalidad asociada a este cáncer biliar. La supervivencia media de los pacientes con CCA es de aproximadamente 6 meses tras el diagnóstico, con una supervivencia a 5 años del 7-20%. En este sentido, es importante destacar que el 2% de los casos mundiales de muerte relacionados con el cáncer corresponden a pacientes con CCA.[3,5]

Dependiendo de su localización anatómica de origen, los CCAs se clasifican en intrahepáticos (iCCA), perihiliares (pCCA) o distales (dCCA).[3,6] El iCCA es el subtipo menos frecuente, representando entre el 10-20% de los casos y puede surgir en cualquier punto del epitelio biliar intrahepático, desde las ramas más pequeñas de los conductos biliares hasta los conductos biliares de segundo orden, también conocidos como conductos biliares segmentarios. El pCCA, anteriormente conocido como tumor de Klatskin, es el subtipo de CCA más común, representando entre el 50-60% de los casos, y pudiéndose desarrollar entre los conductos biliares hepáticos izquierdo y/o derecho hasta la inserción del conducto cístico en el conducto biliar común. Por último, el dCCA, que representan en torno al 20-30% de los casos, puede surgir entre el conducto biliar común por debajo del conducto cístico y el ámpula de Vater, lugar donde el conducto biliar conecta con el conducto pancreático. Con respecto al estadío tumoral, los CCAs se clasifican conforme a la guía clínica del *American Joint of Committee on Cancer* (AJCC), que clasifica los tumores en base al sistema TNM, donde T representa

el tamaño tumoral, N la afectación de los nódulos linfáticos periféricos y M la presencia de metástasis.[7] Los estadíos tumorales se escalan de I al IV, donde los números más bajos representa un estadío tumoral inicial y los números más elevados una situación más avanzada del cáncer.

En la actualidad, se desconoce la etiología de la mayor parte de los CCAs. De hecho, en los países occidentales, alrededor del 50% de los casos son identificados en pacientes carentes de ninguna causa asociada. Sin embargo, una serie de estudios epidemiológicos han descrito ciertos factores de riesgo que incrementan el riesgo de colangiocarcinogénesis.[4,8]

Entre los factores que incrementan de manera elevada el riesgo de desarrollar CCA se encuentran: la presencia de quistes en el colédoco, cálculos biliares, cirrosis, patologías biliares como la enfermedad de Caroli o la colangitis esclerosante primaria (CEP), así como infecciones hepáticas crónicas debido a infección con parásitos hepáticos (*e.g., Opistorchis viverrini* o *Clonorchis sinensis*) o a virus de la hepatitis B (HBV) o C (HCV). En particular, la CEP representa una enfermedad hepática colestásica crónica asociada a procesos autoinmunes frente al epitelio biliar, los cuales pueden provocar inflamación y fibrosis en los conductos, dando lugar a la formación de estenosis biliares multifocales que posteriormente obstruyen las vías biliares tanto intra- como extra-hepáticas. Estos pacientes presentan un riesgo de desarrollar CCA 400 veces mayor que la población general, y en torno al 10-20% de los pacientes con CEP desarrollan cáncer biliar a lo largo de su vida.[9] Asimismo, las infecciones provocadas por los trematodos hepáticos *Opistorchis viverrini* y *Clonorchis sinensis* son altamente prevalentes y representan la mayor causa de CCA en regiones del sudeste asiático tales como Corea, China, Taiwan y Vietnam.[10] Por otro lado, entre aquellos factores que inducen bajo riesgo de desarrollo del CCA pero que son altamente prevalentes a nivel mundial destacan la enfermedad hepática alcohólica, la diabetes tipo II, el tabaco y la enfermedad de hígado graso no alcohólico.

El diagnóstico del CCA se realiza habitualmente mediante la combinación de técnicas de imagen, el análisis de marcadores tumorales no específicos en el suero incluyendo el antígeno carbohidrato 19-9 (CA19-9) o el antígeno carcinoembrionario (CEA), y el estudio histopatológico del tumor, necesario para confirmar el diagnóstico.[11] Las técnicas de imagen se basan generalmente en tomografía computarizada (TC), resonancia magnética (RM), colangiopancreatografía por resonancia magnética (CPRM), tomografía por emisión de positrones (PET), colangiografía transhepática percutánea (PTC), colangiopancreatografía retrógrada endoscópica (CPRE) o ecografía

endoscópica (EUS), cuyo uso habitual depende de la ubicación del tumor. La precisión diagnóstica de los métodos de imagen está influenciada por la localización anatómica y los patrones de crecimiento de los CCAs. La medición de los marcadores tumorales CA19-9 y/o CEA en suero es la única herramienta de biopsia líquida que se usa actualmente en clínica para ayudar en el diagnóstico del CCA, pero su capacidad diagnóstica intrínseca resulta ser escasa, especialmente en etapas iniciales del tumor. Tanto el CA19-9 como el CEA presentan baja sensibilidad y especificidad. Además, altos niveles de CA19-9 en sangre también pueden ser comúnmente detectados en enfermedades biliares benignas, y debido a la deficiencia de actividad de FUT3 hasta el 7% de la población general no expresa el epítopo CA19-9, limitando aún más su uso como estrategia diagnóstica.[12,13] Los métodos de imagen y los marcadores tumorales actuales son poco precisos para el diagnóstico del CCA y el estudio histopatológico presenta limitaciones debido a su carácter invasivo. De hecho, hay veces que se desaconsejan los procedimientos de obtención de biopsia o cepillado tumoral debido a la fragilidad de los pacientes, estadiajes avanzados, localizaciones tumorales que impidan la obtención de cantidad suficiente para el análisis o al posible riesgo de sembrado peritoneal.[11]

El diagnóstico precoz del CCA en pacientes con CEP mediante métodos no invasivos se presenta como un desafío especialmente complejo, ya que frecuentemente las estenosis biliares inflamatorias benignas y malignas detectadas en los pacientes con CEP presentan características morfológicas muy similares, lo que impide mediante las técnicas de imagen demostrar la posible naturaleza neoplásica de la estenosis. Además, el uso del CA19-9 como biomarcador no invasivo para el diagnóstico del CCA en la CEP es limitado, ya que también suele estar elevado en casos de colangitis y colestasis benignas.[14]

Todas estas evidencias resaltan la necesidad de determinar biomarcadores precisos no invasivos para, por un lado, desarrollar futuras estrategias de cribado que permitan un diagnóstico precoz del CCA en poblaciones de alto riesgo (ej., pacientes con CEP) y, por otro lado, para proporcionar un diagnóstico certero no invasivo, lo que ayudaría a disminuir en última instancia la mortalidad relacionada con dicha malignidad.

En este sentido, en los últimos años, las vesículas extracelulares (VEs) se han postulado como contenedores de biomarcadores para diversas patologías.[15,16] Las VEs son esferas membranosas heterogéneas de tamaño nanométrico (30nm-2μM) secretadas por diversos tipos celulares al medio extracelular mediante mecanismos moleculares complejos y que están presentes en fluidos biológicos tales como sangre, orina, bilis,

saliva, etc. En cuanto a su composición molecular, las VEs contienen proteínas, metabolitos, ácidos nucleicos y lípidos, cuya composición y cantidad dependen de la célula de origen. Estas moléculas juegan un papel clave como mediadores intercelulares y pueden ser utilizados como biomarcadores para distintas enfermedades. Además del tipo celular, la carga molecular de las VEs también varía en base a las condiciones fisiopatológicas de cada célula, afectando directamente tanto al destino de las EVs como al efecto inducido por estas vesículas membranosas.

En relación a su biogénesis, las VEs se clasifican en exosomas, microvesículas y cuerpos apoptóticos. Los exosomas presentan una morfología redondeada y un tamaño de aproximadamente 30-200nm. Se forman como vesículas intraluminales (ILV) por la invaginación de la membrana endosomal durante la maduración de los endosomas multivesiculares (MVE) y son secretados al espacio extracelular mediante la fusión de los endosomas tardíos también conocidos como cuerpos multivesiculares (MVB) con la membrana plasmática. Las microvesículas presentan un tamaño de en torno a 50-1000nm y se forman directamente en la membrana plasmática celular por evaginación de la misma para ser liberados al medio extracelular. Por último, los cuerpos apoptóticos son VEs de 800-5000 nm que se liberan durante el proceso de muerte celular programada también conocida como apoptosis.

A pesar de su particular biogénesis, una vez secretados al medio extracelular y alcanzados los fluidos biológicos, tanto los exosomas como las microvesículas muestran una apariencia similar, se superponen en tamaño y, a menudo, presentan una composición molecular común. Por tanto, es difícil determinar el origen de las VEs cuando se aíslan de diversos fluidos biológicos, con lo que el término genérico VEs es el más ampliamente empleado para denominar estas entidades membranosas.

HIPÓTESIS Y OBJETIVOS

La falta de métodos no invasivos y precisos para el diagnóstico del CCA, particularmente importantes para individuos de alto riesgo, hace que en muchos casos estos tumores se diagnostiquen en fases avanzadas, limitando el acceso a posibles opciones curativas como la cirugía. Por ello, y teniendo en cuenta el potencial de las VEs como contenedores de biomarcadores no invasivos de enfermedades y que están presentes en los distintos fluidos biológicos, en este proyecto de tesis doctoral se planteó la hipótesis de que las VEs podrían ser una nueva forma de biopsia líquida que aporte

biomarcadores precisos y no invasivos para la predicción del desarrollo, el diagnóstico precoz y la estimación del pronóstico del CCA.

Por ello, se plantearon los siguientes objetivos:

I. Aislamiento y caracterización de VEs de suero y orina de pacientes con CCA, CEP y colitis ulcerosa (CU), así como de individuos sanos.

II. Determinación del perfil transcriptómico y proteómico de las VEs mediante tecnologías de alto rendimiento, incluyendo *microarrays* de expresión génica y cromatografía líquida de alta resolución acoplada a espectrometría de masas, respectivamente.

III. Identificación de biomarcadores no invasivos en VEs de suero y orina para la predicción de desarrollo, diagnóstico precoz y/o estimación del pronóstico del CCA.

IV. Asociación entre niveles de biomoléculas en VEs de fluidos biológicos con su expresión en hígado humano, tejido tumoral de CCA y subtipos de células que componen el tumor, abordando el concepto de biopsia líquida.

BIOMARCADORES DE ARN EN VEs DE SUERO Y ORINA

En este proyecto, uno de nuestros objetivos fue caracterizar el perfil transcriptómico de VEs de suero y orina de pacientes con CCA, CEP, CU e individuos sanos, con el fin de determinar biomarcadores diagnósticos para el CCA que reflejasen las alteraciones de expresión en el tumor (biopsia líquida).

Para ello, se aislaron VEs de suero y orina de pacientes con CCA [n=12 suero; n=23 orina], CEP [n=6 suero; n=5 orina], CU [n=8 suero; n=12 orina] e individuos sanos [n=9 suero; n=5 orina] a partir de 1 mL de suero y 50 mL de orina mediante un protocolo basado en ultracentrifugaciones diferenciales seriadas. Tras aislar las VEs, se caracterizó la fracción vesicular a nivel morfológico por microscopía electrónica de transmisión (MET), determinación de tamaño y concentración mediante el análisis de seguimiento de nanopartículas (NTA, *Nanoparticle Tracking Analysis*) y el análisis por inmunoblot de los marcadores característicos de las VEs CD63 y CD81. Los resultados mostraron una morfología redondeada, un tamaño en torno a 180 nm y enriquecimiento

de los marcadores CD63 y CD8, indicando la presencia mayoritaria de exosomas y/o microvesículas de pequeño tamaño en las fracciones de VEs aisladas de ambos fluidos biológicos y también su pureza.

A continuación, se analizó el perfil transcriptómico de estas VEs mediante *microarrays* de expresión génica (*Illumina*) que contenía sondas para la detección tanto de ARN mensajero (ARNm) como de ARN no codificante (ARNnc). Se identificaron moléculas de ARNm y ARNnc (*e.g.* microARN, lncARN y snoARN, entre otros) en ambos fluidos biológicos y en los cuatro grupos experimentales. Además, se encontraron perfiles diferenciales de ARN en VEs de suero y orina en pacientes con CCA en comparación con los grupos control (individuos sanos y pacientes con CEP o CU), mostrando alta capacidad diagnóstica en muchos de los ARN. En cuanto a las VEs de suero, se detectó una abundancia diferencial en 2807 transcritos entre los pacientes con CCA y el grupo control formado por pacientes con CEP, CU e individuos sanos; destacando un aumento de los ARNm *RFFL*, *ZNF* y *OR4F3* y de los ARNnc *MIR551B, PMS2L4* y *LOC643955* en pacientes con CCA con respecto al grupo control, con precisión diagnóstica de hasta el 100%, como es el caso de *RFFL* (AUC = 1,000). Con respecto a las VEs de orina, los niveles de 1329 transcritos se encontraron diferencialmente alterados entre los pacientes con CCA y el grupo control formado por pacientes con CEP, CU e individuos sanos, destacando un aumento en los niveles de ARNm de *MT1F*, *GPX3* y *LDHA*, así como de los ARNnc *RNU11* y *LOC257358*, con valor diagnóstico de AUC hasta 0,915 como es el caso de *MT1F*.

A continuación, se correlacionaron los niveles de ARN de los biomarcadores candidatos en fluidos biológicos con su expresión en tumores de CCA y tejido hepático adyacente de dos cohortes de pacientes internacionales independientes [*The cancer genome atlas* (TCGA; n=36 tumor *vs* n=9 adyacente) y la cohorte de Copenhague (n=217 tumor *vs* n=143 adyacente)], así como en colangiocitos humanos tumorales y normales en cultivo, y en VEs secretadas por estas células.

Entre los 2807 transcritos con niveles alterados en VEs de suero de pacientes con CCA, la expresión de 479 ARNm se encontraba alterada en tejido de CCA en comparado con tejido hepático adyacente en ambas dos cohortes internacionales de pacientes y con la misma tendencia observada en el suero. Además, la expresión de 156 ARNm también se encontró alterada en las líneas celulares de CCA EGI1 y TFK1 en comparación con los colangiocitos normales humanos en cultivo. De esos 156 ARNm, 105 mostraban la misma tendencia en VEs secretadas por dichas células. El análisis de ontología génica reveló que estos transcritos participan en procesos fundamentales en la carcinogénesis,

tales como el metabolismo, la comunicación celular y vías de crecimiento celular mediante la modulación de rutas que promueven el cáncer como, por ejemplo, la regulación del ciclo celular, la transición epitelio-mesénquima y la reparación del ADN. Entre estos transcritos, *CMIP*, *NME1* y *GSK1B* fueron los mejores candidatos a biomarcadores de biopsia líquida identificados, ya que además de reflejar los mismos niveles de expresión en el tumor y EVs de sangre, la combinación de estos tres biomarcadores permitió un diagnóstico óptimo del CCA (AUC = 1,00).

Del mismo modo, entre los 1329 transcritos alterados en orina de pacientes con CCA, la expresión de 259 ARNm se encontró alterada en tejido de CCA comparado con tejido hepático adyacente en las cohortes internacionales TCGA y la cohorte de Copenhague, con la misma tendencia que presentaban estos transcritos en VEs de orina. Además, la expresión de 84 ARNm también se encontró alterada en las líneas celulares de CCA EGI1 y TFK1 en comparación con los colangiocitos normales humanos en cultivo, y de ellos, 39 transcritos mostraron la misma tendencia en VEs secretadas por dichas células. En el caso de los biomarcadores de orina, el análisis de ontología génica reveló que estos 39 transcritos estaban principalmente relacionados con procesos de desarrollo y progresión tumoral, tales como la regulación del metabolismo celular, la señalización celular, la transición epitelio-mesénquima y la respuesta inmune. Entre estos ARNm, *UBE2C* y *SERPINB1* se mostraron como mejores biomarcadores de biopsia líquida. La combinación de estos biomarcadores de orina aumentó su capacidad diagnóstica independiente, proporcionando un valor de AUC de 0,812 para el diagnóstico del CCA.

Además del potencial valor diagnóstico de estos biomarcadores de biopsia líquida, y considerando que estos transcritos se encontraron sobre-expresados en tejido tumoral de dos cohortes independientes de pacientes, estos ARNm podrían desempeñar un papel patológico fundamental durante la colangiocarcinogénesis.[17] En este sentido, estudios demuestran que el factor de transcripción *CMIP* está aumentado en tumores de glioma y de cáncer gástrico humano, contribuyendo en el aumento de la proliferación y metástasis tumoral.[18,19] Por otro lado, *CKS1B*, el ARNm que se encontraba en mayor abundancia en VEs de suero de pacientes con CCA, se considera un oncogén. Estudios han descrito un incremento de *CKS1B* en retinoblastoma, carcinoma hepatocelular, carcinoma nasofaríngeo y en mieloma múltiple, y la sobre-expresión de *CKS1B* se ha visto positivamente asociada al incremento en la proliferación, migración, angiogénesis, invasión y quimio-resistencia.[20–24] Así, *UBE2C*, cuyo incremento en la abundancia en VEs de orina de pacientes con CCA ha sido descrita en la presente disertación, ha sido

relacionada en diversos estudios con el cáncer gástrico de tipo intestinal, cáncer de ovario, carcinoma de células escamosas de cabeza y cuello, adenocarcinoma ductal pancreático, carcinoma hepatocelular y cáncer de pulmón, entre otros.[25–31] Por tanto, aunque aún no se ha descrito la función de estos genes en la patogénesis del CCA, estas moléculas podrían estar involucradas en procesos claves de desarrollo o progresión, por lo que futuros estudios de estos genes serán necesarios para determinar su papel en colangiocarcinogénesis.

En conclusión, las VEs de suero y orina de pacientes con CCA contienen perfiles transcriptómicos específicos con alta capacidad diagnóstica. Al menos parte de estos ARN presentes en VEs pueden ser liberados por las células del tejido tumoral y podrían participar en la colangiocarcinogénesis, lo que refleja un nuevo enfoque de las VEs de suero y orina como posible estrategia de biopsia líquida.

BIOMARCADORES PROTEICOS EN VEs DE SUERO

Otro objetivo principal de esta tesis doctoral fue caracterizar el perfil proteómico de las VEs de suero de pacientes con CCA asociado a CEP (CCA-CEP), pacientes con CCA de etiología no CEP, pacientes con CEP que desarrollaron CCA durante el seguimiento, pacientes con CEP aislada que no ha desarrollado CCA e individuos sanos, como estrategia para identificar nuevos biomarcadores precisos y no invasivos para predecir el desarrollo del CCA, diagnosticar estos tumores en estadíos iniciales, y/o estimar el pronóstico de los pacientes con CCA.

Para ello, se aislaron, mediante ultracentrifugación diferencial, VEs de suero de pacientes con CEP aislada (sin desarrollo de CCA en $\geq$5 años tras diagnóstico), con presencia concomitante de CEP y CCA, CEP que desarrollan CCA en el seguimiento clínico (<3 años tras diagnóstico de CEP), así como de individuos sanos para su posterior caracterización a nivel proteico mediante proteómica basada en espectrometría de masas. Conforme a los requerimientos de la Sociedad Internacional de Vesículas Extracelulares (*ISEV, International Society of Extracellular Vesicles*), las vesículas aisladas por ultracentrifugación diferencial seriada fueron caracterizadas por microscopía electrónica, NTA y el análisis de los marcadores de VEs CD63 y CD81 por *inmunoblot*, mostrando una morfología redondeada, enriquecimiento de los marcadores CD63 y CD81, así como un tamaño medio de en torno a 200 nm, lo cual revela que los componentes principales de la fracción de VEs aisladas son exosomas y/o microvesículas de pequeño tamaño. Los resultados proteómicos mostraron una

abundancia diferencial de ciertas proteínas en VEs aisladas de pacientes con CCA-CEP con respecto a los grupos control (*i.e.*, CEP aislada e individuos sanos) con alta capacidad diagnóstica, y de hasta un valor de AUC de 0,931 como es el caso de FRIL. Interesantemente, algunos de los biomarcadores diagnósticos identificados ya se encontraban ya alterados en VEs de pacientes con CEP sin evidencias clínicas de la presencia de CCA pero que desarrollaron dicho cáncer durante el seguimiento. En concreto, la abundancia de FGL1, IGHA1, FIBB o FIBG, entre otros, se encontraba alterada en VEs de suero antes de la detección clínica del CCA, mostrando valores de AUC de hasta 0,900 tras comparar pacientes con CEP que desarrollan CCA en el seguimiento con aquellos pacientes con CEP que no han desarrollado CCA en más de 5 años tras el diagnóstico de CEP.

Con el fin de evaluar si los biomarcadores candidatos para el diagnóstico de CCA eran dependientes o no de la presencia de CEP o por el contrario comunes a otras etiologías, los diferentes biomarcadores identificados se compararon con una cohorte de pacientes que, además de incluir individuos sanos y pacientes con CEP aislada, contenía un grupo de pacientes con CCA de etiología no CEP. En total, se evaluaron los siguientes grupos: CCA-CEP (n=14), CCA sin etiología de CEP (n=26), pacientes con CEP aislada sin desarrollo de CCA tras $\geq$5 años (n=39), pacientes con CEP sin evidencias clínicas de malignidad en el momento de la toma de muestra (sangre) pero que desarrollaron CCA durante el seguimiento (<3 años) (n=10), e individuos sanos (n=41).

Tras el análisis, se identificaron una serie de proteínas específicas para el diagnóstico del CCA en pacientes con CEP, para el diagnóstico del CCA en pacientes sin etiología de CEP y un conjunto de biomarcadores diagnósticos de CCA independientes a su etiología. Los niveles de 14 proteínas permitieron un diagnóstico diferencial de pacientes con CCA-CEP en comparación con pacientes con CEP aislada, todos ellos mostrando una mayor capacidad diagnostica que la del CA19-9 (AUC=0,675), destacando particularmente la alta capacidad diagnóstica de KAIN, HEMO y OIT3 (AUC 0,846; 0,815 y 0,788; respectivamente). Además, la abundancia de 6 proteínas permitió un diagnóstico específico para los CCAs sin etiología de CEP, con valores de AUC de hasta 0,874 en comparación con individuos sanos, como es el caso de COF1. Por otro lado, la abundancia de 23 proteínas permitía el diagnóstico de CCA independientemente de su etiología, entre los que destacan PIGR (AUC=0,812), FIBG (AUC=0,803) y APOA1 (AUC=0,787). La mayoría de los candidatos a biomarcadores mantuvieron su precisión diagnóstica al considerar solo los pacientes con estadíos iniciales del tumor, lo que refuerza su utilidad para un diagnóstico temprano del CCA. Además, los niveles de casi

todos los candidatos a biomarcadores eran independientes al sexo biológico, edad y subtipo de CCA.

A través de la combinación de los biomarcadores cuyos niveles eran independientes a las variables de edad, sexo y subtipo de CCA, y mediante un modelado de aprendizaje automático *logit*, se desarrollaron algoritmos que mostraban un diagnóstico óptimo para la identificación del CCA en pacientes con CEP y otros modelos de regresión para el diagnóstico de CCA en pacientes sin CEP. El modelo de regresión logística binaria que combina 8 biomarcadores proteicos (HV308 / KAIN / HEMO / A2MG / VWF / APOA1 / TLN1 / IMA8) proporcionó la máxima precisión para discriminar pacientes con CCA-CEP concomitante en comparación con pacientes con CEP aislada, manteniendo el 100% en sensibilidad y especificidad en estadíos iniciales de CCA. Además, un algoritmo combinado con PLCH1 y FGL1 mostró un valor diagnóstico de AUC=0,903 y *odds ratio* (OR)=28,8 para pacientes con CCA-CEP en estadíos iniciales comparando con los pacientes con CEP aislada, valores muy superiores al CA19-9 (AUC=0,608, OR=2). Por otro lado, cuando la ausencia de CEP estaba clínicamente confirmada, un modelo *logit* que combina 5 biomarcadores (SAMP / ACTC / RELN / APOA1 / KLKB1) demostró una sensibilidad y especificidad del 100% para discriminar pacientes con CCA en comparación con individuos sanos, y también en estadíos iniciales del tumor. Asimismo, un algoritmo que combina SAMP y A1AT permitió diagnosticar el CCA en fases iniciales en pacientes sin CEP en comparación con individuos sanos (AUC=0,863, OR=18,5).

De cara a entender el posible origen de los biomarcadores presentes en VEs circulantes de pacientes con CCA, se analizó la expresión de dichas proteínas en bases de datos con expresión génica de 61 tejidos/órganos humanos, en las diferentes poblaciones celulares del hígado mediante secuenciación de célula única (en ingles, *single-cell RNA sequencing,* o scRNA-seq), así como en scRNA-seq de tejido tumoral de CCA. Los resultados transcriptómicos de múltiples órganos revelaron que casi todos los biomarcadores de VEs se expresaban principalmente en tejidos hepatobiliares. El análisis de scRNA-seq de hígado humano mostró que la expresión de dichos biomarcadores candidatos era en algunos casos específica de ciertos tipos celulares, mientras que en otros casos su expresión es común en diferentes tipos de células hepáticas. Por otro lado, el análisis de scRNA-seq de tumores de CCA indicó que algunos biomarcadores, incluidos *PIGR, FGG, SERPINA1* y *FGL1*, se expresan principalmente en colangiocitos tumorales, mientras que la expresión de *VWF* y *ALPL* se encuentra principalmente en células endoteliales.

A continuación, con el objetivo de determinar biomarcadores proteicos en VEs de suero con capacidad de predecir la supervivencia de los pacientes con CCA, se realizó un estudio multivariable de regresión de Cox considerando variables como la edad, sexo, estadío tumoral, subtipo de CCA, presencia de CEP, cirrosis, y valores séricos de CA19-9, bilirrubina, fosfatasa alcalina, ALT, AST y GGT. Tras este análisis, se identificó que los niveles de los biomarcadores FCN2, MMRN1, SDPR, ACTC, ADIPO, CO4A, FA9, PSB3 y PSA2 en VEs circulantes eran predictores pronósticos independientes de dichas variables anteriormente mencionadas. Además, un panel compuesto por los biomarcadores FCN2, SDPR y FA9 se vio fuertemente asociado con la supervivencia global de los pacientes. De hecho, los pacientes que presentaban altos niveles en estos tres biomarcadores mostraron una supervivencia mayor a 10 años, mientras que los pacientes con bajos niveles en 2 o 3 de estos biomarcadores tenían una supervivencia media menor a 7 meses.

En conclusión, este estudio ha demostrado la capacidad de las VEs de suero de contener biomarcadores proteicos para la predicción del desarrollo de CCA, así como para la detección precoz de tumores en individuos con etiología de CEP, en individuos sin CEP, así como para la detección del CCA independientemente de la etiología de la enfermedad, superando el estándar actual de CA19-9. Este estudio refuerza la idea de que los CCAs que surgen de diferentes etiologías pueden contener proteínas en VEs de suero comunes y diferentes, y, por tanto, refleja la necesidad de utilizar biomarcadores o algoritmos combinatorios adecuados y bien definidos para identificar el CCA en subgrupos de pacientes específicos, que permitan realizar un diagnóstico personalizado de acuerdo con las características clínicas de cada paciente. El hecho de que la mayoría de los biomarcadores circulantes se exprese en tumores de CCA y preferentemente en colangiocitos malignos refuerza el uso de las VEs como estrategia innovadora de biopsia líquida derivada de células tumorales. Finalmente, la asociación de los niveles de ciertas proteínas con la supervivencia del paciente muestra que las VEs de suero también contienen biomarcadores proteicos con alta capacidad pronóstica para los individuos con CCA.

DISCUSIÓN GENERAL

El CCA representa un problema clínico, social y económico a nivel mundial. Aunque actualmente se considera una enfermedad rara, su incidencia y mortalidad asociada están aumentando de manera global. La escasa precisión de las técnicas de imagen actuales y la falta de sensibilidad y especificidad de marcadores tumorales circulantes

empleados para el diagnóstico del CCA conllevan la necesidad de utilizar pruebas invasivas (biopsia o citología) para confirmar su diagnóstico, imposibilitando también el desarrollo de programas de cribado no invasivo y fiables del cáncer en individuos de alto riesgo. Por ello, es importante desarrollar herramientas seguras y fiables para la predicción de desarrollo y detección temprana de tumores especialmente en individuos con factores de riesgo (*e.g.,* CEP), lo cual permitirá detectar los tumores en fases más tempranas y recibir tratamientos potencialmente curativos (quirúrgicos).

En este sentido, la biopsia líquida ofrece la posibilidad de conocer el estado de salud de ciertos tejidos específicos del cuerpo mediante la detección de moléculas liberadas por los tejidos en fluidos biológicos de fácil acceso, representando una alternativa segura al procedimiento invasivo que implica la biopsia.[32] El gran desafío de la biopsia líquida consiste en que las moléculas liberadas a biofluidos por tejidos afectados se encuentran en cantidades relativamente bajas, por lo que identificarlas, cuantificarlas y asociarlas al órgano o tejido particular donde se desarrolla el proceso patológico requiere un gran esfuerzo tecnológico. Además, para poder determinar el posible origen de esas moléculas liberadas a los fluidos biológicos hay que considerar que los órganos y tejidos representan típicamente masas diversas compuestas por distintos tipos celulares que interactúan entre ellas y pueden diferir ampliamente en la ontogenia, función y expresión génica. En este sentido, la tecnología de secuenciación scRNA-seq desarrollada recientemente es una herramienta valiosa para diseccionar la heterogeneidad celular en sistemas complejos y proporciona información sobre las células individuales que componen tejidos y órganos.[33]

Entre los fluidos empleados para biopsia líquida, el suero y la orina son los más utilizados debido a que permiten una extracción sencilla de la muestra y proporcionan suficiente volumen para los análisis requeridos. Estos fluidos, contienen una gran variedad de moléculas entre las que se encuentran las VEs. El análisis de VEs de suero y de orina mediante técnicas ómicas de alto rendimiento proporcionan una serie de ventajas con respecto al fluido total. Por un lado, en proteómica basada en espectrometría de masas las proteínas más abundantes tienden a enmascarar la detección de proteínas menos abundantes, que podrían ser biomarcadores potenciales. Por ello, el aislamiento de VEs permite separar proteínas altamente abundantes y, mediante el análisis de ese sub-proteoma, aumentar la capacidad de identificación y detección de proteínas que se encuentran en menor concentración en fluidos biológicos complejos como el suero.[34] En segundo lugar, para que una molécula pueda ser candidata a biopsia líquida, debe ser secretada por el tejido de origen y localizada en fluidos circulantes. Debido al conocimiento de la biogénesis de las VEs, se sabe que

estas vesículas se generan en las células, son secretadas al medio extracelular y alcanzan diversos fluidos biológicos.[35] Por tanto, moléculas identificadas en estas VEs podrían ser buenos biomarcadores de biopsia líquida y reflejar el estado celular de donde se originan. Por último, la encapsulación de los ácidos ribonucleicos en VEs previene su fragmentación o degradación por parte de las nucleasas en circulación, convirtiéndolos en candidatos más estables.

Sin embargo, aunque las VEs ofrecen una estrategia prometedora para la búsqueda de biomarcadores de biopsia líquida, su traslacionalidad clínica se encuentra limitada, debido principalmente a la complejidad, heterogeneidad y variabilidad de las técnicas de aislamiento que provocan una falta de reproducibilidad de los resultados. Por lo tanto, la consideración del uso de plataformas proteómicas y transcriptómicas de alto rendimiento para la identificación de posibles biomarcadores tempranos no invasivos en VEs y la posterior validación de dichos biomarcadores en fluidos no procesados a través de técnicas sencillas y reproducibles representa una estrategia racional para el descubrimiento y aplicación clínica de biomarcadores no invasivos.

En este sentido, la reacción cuantitativa en cadena de la polimerasa (qPCR) y los ensayos de proteínas basados en inmuno-afinidad como el ensayo por inmuno-absorción ligado a enzimas (ELISA) se han postulado como las plataformas analíticas preferidas por la comunidad científica para la validación de biomarcadores de ARN y proteínas en fluidos, respectivamente, debido a que son técnicas reproducibles, rápidas y sin gran demanda de equipamiento que permiten analizar un gran número de muestras de manera simultánea.

En el futuro, la demanda clínica del aumento de precisión de la biopsia líquida en clínica y la atención individualizada será cubierta por algoritmos de aprendizaje automáticos que combinen una serie de biomarcadores individuales.[36] A menudo, los biomarcadores individuales carecen de precisión, debido a su variabilidad de expresión tanto en individuos sanos como en enfermos, y a la naturaleza heterogénea de enfermedades como el CCA. Sin embargo, los algoritmos construidos automáticamente a partir de múltiples biomarcadores superan la sensibilidad y especificidad de los biomarcadores individuales y mejoran su rendimiento diagnóstico.

La validación futura de los biomarcadores de CCA propuestos en este estudio y otros que surjan en estudios en curso, junto con el desarrollo de modelos logísticos que combinen diferentes tipos de biomoléculas podrán permitir mejorar la detección precoz del CCA, aumentando sustancialmente la posibilidad de utilizar tratamientos más

eficientes para los pacientes y, directamente mejorando el bienestar social y minimizando los costes sanitarios.

Este beneficio global solo será posible mediante la cooperación interdisciplinaria entre médicos, investigadores básicos y traslacionales, biólogos computacionales, bioestadísticos y epidemiólogos, a través del desarrollo de consorcios colaborativos internacionales como la Red Europea para el Estudio del Colangiocarcionoma (ENSCCA), el establecimiento de bases de datos colaborativas y accesibles, el uso de registros internacionales de pacientes, una gestión óptima de biobancos nacionales e internacionales y una concienciación activa de los pacientes para lograr su implicación en la investigación mediante la donación de muestras. Así, se podrán identificar nuevos biomarcadores no invasivos precisos para el diagnóstico temprano del CCA, posibilitando un abordaje terapéutico potencialmente curativo que resultará ciertamente en una mejoría del pronóstico y de la calidad de vida de los pacientes.

CONCLUSIONES

1. Las VEs de suero y orina aisladas mediante ultracentrifugación diferencial seriada pertenecen principalmente a exosomas y microvesículas, caracterizadas por su morfología típica redondeada, un tamaño en torno a 200 nm y un enriquecimiento en los marcadores característicos de VEs, tales como CD63 y CD81.

2. El perfil transciptómico y proteómico de las VEs reveló que, entre todas las moléculas de ARN y proteínas identificadas, se detectaron algunas biomoléculas específicas cuyos niveles estaban alterados en pacientes con CCA en comparación con individuos sanos, pacientes con CU y pacientes con CEP aislada. El perfil proteómico también mostro diferencias entre pacientes con CCA en un contexto de CEP y pacientes con CCA de etiología no CEP, así como entre pacientes con CEP que desarrollan CCA y los que no desarrollan cáncer biliar.

3. El análisis estadístico de la capacidad diagnóstica o pronóstica de las moléculas de ARN y/o proteínas reveló que las VEs contienen biomarcadores precisos para el diagnóstico de los CCAs, con una sensibilidad y especificidad mayor al marcador tumoral no invasivo actual CA19-9. Los biomarcadores de VEs revelaron capacidad de predicción del desarrollo de CCA en pacientes con CEP, de detección del tumor en estadíos iniciales (tanto en pacientes con etiología de CEP como en pacientes de etiología no CEP), y de estimación del pronóstico de los individuos con CCA.

4. Ciertos biomarcadores de biopsia líquida se expresan en tumores de CCA, preferentemente en los colangiocitos malignos. Este hecho confirma la posibilidad de que los biomarcadores de VEs descritos en este trabajo sean producidos por las propias células de CCA y liberados al suero y, en algunos casos, posteriormente a la orina, reflejando el contenido del tumor de manera no invasiva. Además, estos resultados muestran que además de biomarcadores, estas moléculas podrían ser nuevas posibles dianas terapéuticas ya que podrían estar implicadas en la patogénesis de la enfermedad.

5. Los algoritmos de aprendizaje automático que combinan biomarcadores parecen ser más precisos, en términos diagnósticos y/o pronósticos, que dichos biomarcadores de forma individual, tanto en sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, precisión, razón de probabilidades y razón de riesgo. Estos algoritmos de aprendizaje automático pueden permitir la predicción, el diagnóstico o el pronóstico correctos de una población más amplia, disminuyendo los resultados con falsos positivos o negativos que puedan surgir en las pruebas de biomarcadores individuales.

6. El estudio de transcriptómica mostró que un modelo de regresión logística compuesto por los ARNm *CMIP*, *NME1* y *CKS1B* en VEs de suero discrimina con el 100% de precisión pacientes con CCA de pacientes con CEP, CU o individuos sanos.

7. En pacientes con CEP, el algoritmo combinatorio de los niveles de las proteínas HV308, KAIN, HEMO, A2MG, VWF, APOA1, TLN1 e IMA8 en suero permitieron el diagnóstico de CCA en pacientes con CEP en el 100% de los casos, y el modelo logístico incluyendo únicamente las proteínas PLCH1 y FGL1 ofrecieron una precisión del 91% para la detección temprana del CCA en pacientes con CEP

8. En pacientes sin CEP, la combinación de las proteínas SAMP, ACTC, RELN, APOA1 y KLKB1 ofrecían un 100% de sensibilidad y especificidad para el diagnóstico del CCA en comparación con los individuos sanos, mientras que la inclusión algorítmica de únicamente las proteínas SAMP y A1AT ofreció un diagnóstico de CCA en comparación con individuos sanos con 77% de sensibilidad y 90% de especificidad.

REFERENCIAS

1. Banales JM, Cardinale V, Carpino G, et al. Expert consensus document: Cholangiocarcinoma: current knowledge and future perspectives consensus statement from the European Network for the Study of Cholangiocarcinoma (ENS-CCA). *Nat Rev Gastroenterol Hepatol*. 2016;13(5):261-280. doi:10.1038/NRGASTRO.2016.51

2. Brindley PJ, Bachini M, Ilyas SI, et al. Cholangiocarcinoma. *Nat Rev Dis Prim*. 2021;7(1). doi:10.1038/S41572-021-00300-2

3. Banales JM, Marin JJG, Lamarca A, et al. Cholangiocarcinoma 2020: the next horizon in mechanisms and management. *Nat Rev Gastroenterol Hepatol*. 2020;17(9):557-588. doi:10.1038/S41575-020-0310-Z

4. Khan SA, Tavolari S, Brandi G. Cholangiocarcinoma: Epidemiology and risk factors. *Liver Int*. 2019;39 Suppl 1(S1):19-31. doi:10.1111/LIV.14095

5. Bertuccio P, Malvezzi M, Carioli G, et al. Global trends in mortality from intrahepatic and extrahepatic cholangiocarcinoma. *J Hepatol*. 2019;71(1):104-114. doi:10.1016/J.JHEP.2019.03.013

6. Blechacz B, Komuta M, Roskams T, Gores GJ. Clinical diagnosis and staging of cholangiocarcinoma. *Nat Rev Gastroenterol Hepatol*. 2011;8(9):512-522. doi:10.1038/NRGASTRO.2011.131

7. AJCC Cancer Staging Manual. *AJCC Cancer Staging Man*. Published online 2017. doi:10.1007/978-3-319-40618-3

8. Rodrigues PM, Olaizola P, Paiva NA, et al. Pathogenesis of Cholangiocarcinoma. *Annu Rev Pathol*. 2021;16:433-463. doi:10.1146/ANNUREV-PATHOL-030220-020455

9. Boonstra K, Weersma RK, van Erpecum KJ, et al. Population-based epidemiology, malignancy risk, and outcome of primary sclerosing cholangitis. *Hepatology*. 2013;58(6):2045-2055. doi:10.1002/HEP.26565

10. Kim TS, Pak JH, Kim JB, Bahk YY. Clonorchis sinensis, an oriental liver fluke, as a human biological agent of cholangiocarcinoma: a brief review. *BMB Rep*. 2016;49(11):590-597. doi:10.5483/BMBREP.2016.49.11.109

11. Forner A, Vidili G, Rengo M, Bujanda L, Ponz-Sarvisé M, Lamarca A. Clinical presentation, diagnosis and staging of cholangiocarcinoma. *Liver Int*. 2019;39 Suppl 1(S1):98-107. doi:10.1111/LIV.14086

12. Onal C, Colakoglu T, Ulusan SN, Yapar AF, Kayaselcuk F. Biliary obstruction induces extremely elevated serum CA 19-9 levels: case report. *Onkologie*. 2012;35(12):780-782. doi:10.1159/000345110

13. Wannhoff A, Hov JR, Folseraas T, et al. FUT2 and FUT3 genotype determines CA19-9 cut-off values for detection of cholangiocarcinoma in patients with primary sclerosing cholangitis. *J Hepatol*. 2013;59(6):1278-1284. doi:10.1016/J.JHEP.2013.08.005

14. Grimsrud MM, Folseraas T. Pathogenesis, diagnosis and treatment of premalignant and malignant stages of cholangiocarcinoma in primary sclerosing cholangitis. *Liver Int*. 2019;39(12):2230-2237. doi:10.1111/LIV.14180

15. Yáñez-Mó M, Siljander PRM, Andreu Z, et al. Biological properties of extracellular vesicles and their physiological functions. *J Extracell vesicles*. 2015;4(2015):1-60. doi:10.3402/JEV.V4.27066

16. Lapitz A, Arbelaiz A, Olaizola P, et al. Extracellular Vesicles in Hepatobiliary Malignancies. *Front Immunol*. 2018;9(OCT). doi:10.3389/FIMMU.2018.02270

17. Lapitz A, Arbelaiz A, O'Rourke CJ, et al. Patients with Cholangiocarcinoma Present Specific RNA Profiles in Serum and Urine Extracellular Vesicles Mirroring the Tumor Expression: Novel Liquid Biopsy Biomarkers for Disease Diagnosis. *Cells*. 2020;9(3). doi:10.3390/CELLS9030721

18. Zhang J, Huang J, Wang X, et al. CMIP is oncogenic in human gastric cancer cells. *Mol Med Rep*. 2017;16(5):7277-7286. doi:10.3892/MMR.2017.7541

19. Wang B, Wu ZS, Wu Q. CMIP Promotes Proliferation and Metastasis in Human Glioma. *Biomed Res Int*. 2017;2017. doi:10.1155/2017/5340160

20. Zeng Z, Gao ZL, Zhang ZP, et al. Downregulation of CKS1B restrains the proliferation, migration, invasion and angiogenesis of retinoblastoma cells through the MEK/ERK signaling pathway. *Int J Mol Med*. 2019;44(1):103-114. doi:10.3892/IJMM.2019.4183

21. Kang YS, Jeong EJ, Seok HJ, et al. Cks1 regulates human hepatocellular carcinoma cell progression through osteopontin expression. *Biochem Biophys Res Commun*. 2019;508(1):275-281. doi:10.1016/J.BBRC.2018.11.070

22. Xu L, Fan S, Zhao J, et al. Increased expression of Cks1 protein is associated with lymph node metastasis and poor prognosis in nasopharyngeal carcinoma. *Diagn Pathol*. 2017;12(1). doi:10.1186/S13000-016-0589-9

23. Shi L, Wang S, Zangari M, et al. Over-expression of CKS1B activates both MEK/ERK and JAK/STAT3 signaling pathways and promotes myeloma cell drug-resistance. *Oncotarget*. 2010;1(1):22-33. doi:10.18632/ONCOTARGET.105

24. Huang CW, Lin CY, Huang HY, et al. CKS1B overexpression implicates clinical aggressiveness of hepatocellular carcinomas but not p27(Kip1) protein turnover: an independent prognosticator with potential p27 (Kip1)-independent oncogenic attributes? *Ann Surg Oncol*. 2010;17(3):907-922. doi:10.1245/S10434-009-0779-8

25. Zhang J, Liu X, Yu G, et al. UBE2C Is a Potential Biomarker of Intestinal-Type Gastric Cancer With Chromosomal Instability. *Front Pharmacol*. 2018;9(AUG). doi:10.3389/FPHAR.2018.00847

26. Li J, Zhi X, Shen X, et al. Depletion of UBE2C reduces ovarian cancer malignancy and reverses cisplatin resistance via downregulating CDK1. *Biochem Biophys Res Commun*. 2020;523(2):434-440. doi:10.1016/J.BBRC.2019.12.058

27. Jin Z, Zhao X, Cui L, et al. UBE2C promotes the progression of head and neck squamous cell carcinoma. *Biochem Biophys Res Commun*. 2020;523(2):389-397. doi:10.1016/J.BBRC.2019.12.064

28. Wang X, Yin L, Yang L, et al. Silencing ubiquitin-conjugating enzyme 2C inhibits proliferation and epithelial-mesenchymal transition in pancreatic ductal adenocarcinoma. *FEBS J*. 2019;286(24):4889-4909. doi:10.1111/FEBS.15134

29. Wei ZI, Liu YI, Qiao SH, et al. Identification of the potential therapeutic target gene UBE2C in human hepatocellular carcinoma: An investigation based on GEO and TCGA databases. *Oncol Lett*. 2019;17(6):5409-5418. doi:10.3892/OL.2019.10232

30. Xiong Y, Lu J, Fang Q, et al. UBE2C functions as a potential oncogene by enhancing cell proliferation, migration, invasion, and drug resistance in hepatocellular carcinoma cells. *Biosci Rep*. 2019;39(4). doi:10.1042/BSR20182384

31. Wu Y, Jin D, Wang X, et al. UBE2C Induces Cisplatin Resistance via ZEB1/2-Dependent Upregulation of ABCG2 and ERCC1 in NSCLC Cells. *J Oncol*. 2019;2019. doi:10.1155/2019/8607859

32. Landegren U, Hammond M. Cancer diagnostics based on plasma protein biomarkers: hard times but great expectations. *Mol Oncol*. 2021;15(6):1715-1726. doi:10.1002/1878-0261.12809

33. Tanay A, Regev A. Scaling single-cell genomics from phenomenology to mechanism. *Nature*. 2017;541(7637):331-338. doi:10.1038/NATURE21350

34. Burton JB, Carruthers NJ, Stemmer PM. Enriching extracellular vesicles for mass spectrometry. *Mass Spectrom Rev*. Published online 2021. doi:10.1002/MAS.21738

35. Van Niel G, D'Angelo G, Raposo G. Shedding light on the cell biology of extracellular vesicles. *Nat Rev Mol Cell Biol*. 2018;19(4):213-228. doi:10.1038/NRM.2017.125

36. Christou CD, Tsoulfas G. Challenges and opportunities in the application of artificial intelligence in gastroenterology and hepatology. *World J Gastroenterol*. 2021;27(37):6191-6223. doi:10.3748/WJG.V27.I37.6191

LIST OF PUBLICATIONS

1. Santos-Laso A, Izquierdo-Sanchez L, Rodrigues PM, Huang BQ, Azkargorta M, **Lapitz A**, Munoz-Garrido P, Arbelaiz A, Caballero-Camino FJ, Fernández-Barrena MG, Jimenez-Agüero R, Argemi J, Aragon T, Elortza F, Marzioni M, Drenth JPH, LaRusso NF, Bujanda L, Perugorria MJ, Banales JM. Proteostasis disturbances and endoplasmic reticulum stress contribute to polycystic liver disease: New therapeutic targets. Liver Int. 2020 Jul;40(7):1670-1685. doi: 10.1111/liv.14485. Epub 2020 May 6. PMID: 32378324; PMCID: PMC7370945.

2. Macias RIR, Muñoz-Bellvís L, Sánchez-Martín A, Arretxe E, Martínez-Arranz I, **Lapitz A**, Gutiérrez ML, La Casta A, Alonso C, González LM, Avila MA, Martinez-Chantar ML, Castro RE, Bujanda L, Banales JM, Marin JJG. A Novel Serum Metabolomic Profile for the Differential Diagnosis of Distal Cholangiocarcinoma and Pancreatic Ductal Adenocarcinoma. Cancers (Basel). 2020 May 31;12(6):1433. doi: 10.3390/cancers12061433. PMID: 32486461; PMCID: PMC7352809.

3. Hernández A, Arab JP, Reyes D, **Lapitz A**, Moshage H, Bañales JM, Arrese M. Extracellular Vesicles in NAFLD/ALD: From Pathobiology to Therapy. Cells. 2020 Mar 27;9(4):817. doi: 10.3390/cells9040817. PMID: 32231001; PMCID: PMC7226735.

4. **Lapitz A**, Arbelaiz A, O'Rourke CJ, Lavin JL, Casta A, Ibarra C, Jimeno JP, Santos-Laso A, Izquierdo-Sanchez L, Krawczyk M, Perugorria MJ, Jimenez-Aguero R, Sanchez-Campos A, Riaño I, Gónzalez E, Lammert F, Marzioni M, Macias RIR, Marin JJG, Karlsen TH, Bujanda L, Falcón-Pérez JM, Andersen JB, Aransay AM, Rodrigues PM, Banales JM. Patients with Cholangiocarcinoma Present Specific RNA Profiles in Serum and Urine Extracellular Vesicles Mirroring the Tumor Expression: Novel Liquid Biopsy Biomarkers for Disease Diagnosis. Cells. 2020 Mar 14;9(3):721. doi: 10.3390/cells9030721. PMID: 32183400; PMCID: PMC7140677.

5. Rodrigues PM, **Lapitz A**, Simão A, Perugorria MJ, Arrese M, Castro RE, Banales JM. Extracellular Vesicles in Non-alcoholic Fatty Liver Disease: Key Players in Disease Pathogenesis and Promising Biomarker Tools. NAFLD and NASH; Springer Nature Switzerland AG 2020 Feb. Volume I, Chapter 9:157-180. doi: 10.1007/978-3-030-37173-9_9

6. Banales JM, Iñarrairaegui M, Arbelaiz A, Milkiewicz P, Muntané J, Muñoz-Bellvis L, La Casta A, Gonzalez LM, Arretxe E, Alonso C, Martínez-Arranz I, **Lapitz A**, Santos-Laso A, Avila MA, Martínez-Chantar ML, Bujanda L, Marin JJG, Sangro B, Macias RIR. Serum Metabolites as Diagnostic Biomarkers for Cholangiocarcinoma, Hepatocellular Carcinoma, and Primary Sclerosing Cholangitis. Hepatology. 2019 Aug;70(2):547-562. doi: 10.1002/hep.30319. Epub 2019 Feb 14. PMID: 30325540; PMCID: PMC6767196.

7. **Lapitz A**, Arbelaiz A, Olaizola P, Aranburu A, Bujanda L, Perugorria MJ, Banales JM. Extracellular Vesicles in Hepatobiliary Malignancies. Front Immunol. 2018 Oct 12;9:2270. doi: 10.3389/fimmu.2018.02270. PMID: 30369925; PMCID: PMC6194158.

8. Olaizola P, Lee-Law PY, Arbelaiz A, **Lapitz A**, Perugorria MJ, Bujanda L, Banales JM. MicroRNAs and extracellular vesicles in cholangiopathies. Biochim Biophys Acta Mol Basis Dis. 2018 Apr;1864(4 Pt B):1293-1307. doi: 10.1016/j.bbadis.2017.06.026. Epub 2017 Jul 13. PMID: 28711597.

9. Arbelaiz A, Azkargorta M, Krawczyk M, Santos-Laso A, **Lapitz A**, Perugorria MJ, Erice O, Gonzalez E, Jimenez-Agüero R, Lacasta A, Ibarra C, Sanchez-Campos A, Jimeno JP, Lammert F, Milkiewicz P, Marzioni M, Macias RIR, Marin JJG, Patel T, Gores GJ, Martinez I, Elortza F, Falcon-Perez JM, Bujanda L, Banales JM. Serum extracellular vesicles contain protein biomarkers for primary sclerosing cholangitis and cholangiocarcinoma. Hepatology. 2017 Oct;66(4):1125-1143. doi: 10.1002/hep.29291. Epub 2017 Aug 26. PMID: 28555885.

ACKNOWLEDGEMENTS (AGRADECIMIENTOS)