TEAD4 promotes tumorigenesis via regulation of HSP70B expression in hepatocellular carcinoma.

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Nadia Tosti

aus Italien

2020

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Alex Odermatt

Prof. Dr. Luigi Maria Terracciano

PD. Dr. Marianna Kruithof de Julio

Basel, den 21 Mai 2019

Prof. Dr. Martin Spiess

Dekan

- *Success is stumbling from failure to failure*

with no loss of enthusiasm -

[Winston S. Churchill]

TABLE OF CONTENTS

1. SUMMARY

Hepatocellular carcinoma (HCC) represents more than 90% of primary liver cancers with almost 800,000 deaths each year and increasing incidence. Despite the progress done in preventing, treating and improving patients' life, incidence and mortality still rising. Surveillance programmes allow the diagnosis of early stage tumors that can benefit from therapies with curative intent such as resection, liver transplantation or local ablation. On the other hand, advanced HCC stages have limited benefit from chemoembolization and systemic treatments like sorafenib or regorafenib. Recent development of next generations sequence technologies has been useful to unveil the genetic and molecular landscape of HCC. However, the high heterogeneity of HCC together difficulties the development of more effective therapies. Thus, the identification of new molecular target is vital to develop more effective therapies for HCC patients.

TEAD4 is a member of the transcriptional enhancer factor family (TEF) that has been found dysregulated in different tumor entities. Several studies have validated its oncogenic role in the tumorigenic process by regulating key pathways involved in proliferation, migration and invasiveness. However, *TEAD4* role in liver carcinogenesis remains still to be elucidate.

The present work demonstrated that *TEAD4* promote hepatocarcinogenesis by regulating the transcription and the expression of HSP70B, member of the heat shock protein family. The evidence provided here suggest a novel mechanism

inducing hepatocarcinogenesis that be exploit as new potential therapeutic target for HCC treatment.

2. ABBREVIATIONS

AFB1: Aflatoxin B1

AFP: Alpha-fetoprotein

BCLC: Barcelona Clinic Liver Cancer

ChIP: Chromatin immunoprecipitation

CLIP: Cancer of the Liver Italian program

CNA: Copy number alteration

CTGF: Connective tissue growth factor

DAAs: Direct-acting antiviral agents

DBD: DNA binding domain

EMT: epithelial to mesenchymal transition

FDR: False Discovery Rate

HBV: Hepatitis B Virus

HCC: Hepatocellular Carcinoma

HCV: Hepatitis C Virus

LATS1/2: Large tumor suppressor 1/2

LSECs: Liver sinusoidal endothelial cells

LT: Liver transplantation

NAFLD: Non alcoholic fatty liver disease

OS: overall survival

PEI: Percutaneous ethanol injection

PLC: Primary Liver Cancer

RFA: Radiofrequency ablation

SAV: Salvador Homolog 1

SVR: Sustained virologic response

TACE: Transarterial chemoembolization

TERT: Telomerase Reverse Transcriptase

TF: Transcription factor

TNM: Tumor Node Metastasis

TSS: Transcription starting site

VGLL4: Vestigial like family member 4

YBD: YAP binding domain

3. INTRODUCTION

3.1 The human liver

The liver is the second biggest organ in our body, after the skin. The weight oscillates from 1300 to 1700g, based on sex and body mass. It is located in the abdominal cavity, right under the diaphragm and it is divided in four different lobes (left, right, caudate and quadrate).[1] It is known to be a metabolically active organ responsible for three fundamental functions: detoxification, synthesis and storage. The first process is aimed to eliminate toxins from our body. The second function includes the metabolism of proteins, fats, carbohydrates as well as the synthesis of plasma proteins (such as albumin) and the production of the bile.

The liver is a very complex organ characterized by different and peculiar cell types including hepatocytes, biliary epithelial cells, stellate cells, Kupffer cells and liver sinusoidal endothelial cells (Figure 1B).[2] Hepatocytes are considered the working cells of the liver and they are responsible for the majority of the metabolic and regulatory functions of the organ, occupying 80% of its entire mass. The biliary epithelial cells or cholangiocytes, represent the second most frequent cellular component (3-5%), they build the entire bile ducts and they have the important role of modify the composition of the bile when it reaches the bile ducts.[3] The stellate cells represents the minor cell population in the liver. In physiological condition, they have the role of storage of vitamin A in their cytoplasm. Nevertheless, they play a remarkable function during liver injury. They are able to change their phenotypic conformation from “quiescent”, non-proliferating cells, to “activated”, myofibroblast-like cells. This results in the production of several extracellular matrix components

such as the collagen, whose role in the development of hepatic fibrosis and cirrhosis is well known.[4] Another important cellular component is represented by Kupffer cells, considered as the bodyguards of the hepatocytes. Indeed, they are the largest population of resident macrophages in the body and they have a pivotal role in the innate immune response.[5] Finally, the liver sinusoidal endothelial cells (LSECs) function as interface between the blood flow, in particular arterial and portal blood, and hepatocytes and hepatic stellate cells on the other side.[6] Therefore, this permeable barrier allow exchanges of molecules through their fenestrae.

All the hepatic cells are arranged around the lobule, considered the functional unit of the liver (Figure 1A). It is characterized by an hexagonal shape with portal triads (hepatic artery, portal vein and bile duct) at each corner.[2] The vessels of the portal triads send distributing branches along the sides of the lobule, and these branches open into the sinusoids. The long axis of the lobule is transverse by the central vein, and this vessel receives blood from the sinusoids. Interconnecting foils of hepatocytes are disposed in a radial pattern from the central vein to the perimeter of the lobule.

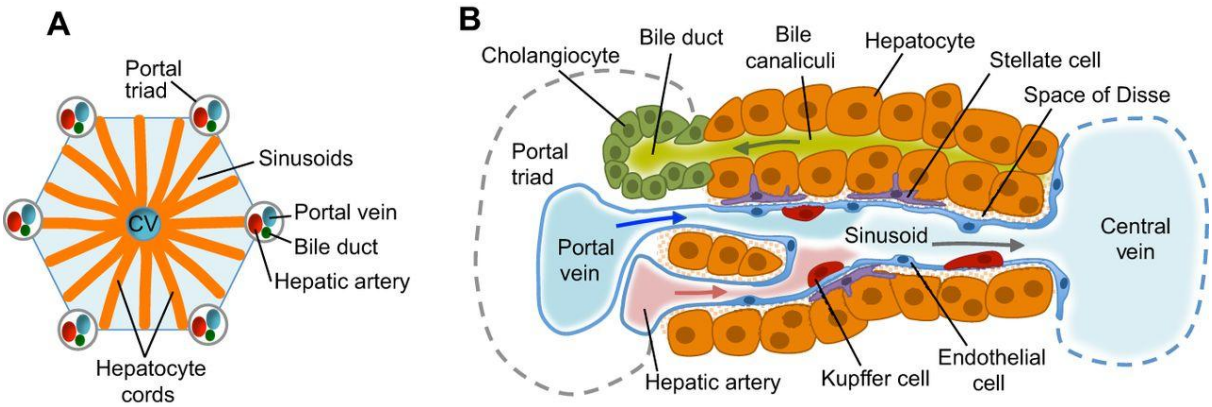


Figure 1: **Structure of the liver.** Taken from Gordillo et al.[7] **A)** Schematic representation of a lobule, the functional unit of the liver. From the central vein (CV), hepatocytes cords spread toward portal triads, composed of portal vein, bile duct and hepatic artery; **B)** organization of the different liver cell types within each lobule.

3.2 Hepatocellular carcinoma

Several disorders can affect the well-being of the liver, including hepatitis, cirrhosis and primary liver cancer (PLC). PLC is responsible for almost 800,000 death each year, being the second leading cause of cancer related death and the fifth most frequent cancer worldwide.[8] Hepatocellular carcinoma (HCC) is considered the most common primary liver cancer, accounting for approximately 80-90% of PLCs. The remaining 6-15% of PLCs includes intrahepatic cholangiocarcinoma and extrahepatic bile-duct carcinoma.[9] The incidence of HCC is two to four times higher in men than women and the majority of them are > 45 years of age reaching a peak at 70 years. Moreover, differences on the incidence of HCC are present between countries. According to the GLOBOCAN database, that collects information about cancer trends and geographic diversity around the globe, the burden of HCC is more localized in countries having developing economies.[10] Indeed, GLOBOCAN 2018 estimates around 13 countries where the incidence of HCC is higher.[11] South-Eastern Asia, Northern and Western Africa and Polynesia show mildly high liver cancer incidence (11-20 cases per 100,000 male inhabitants) or high (> 20 per 100,000). On the contrary, the incidence decreases in Central and Southern America (6.7 and 5.8 per 100,000 male inhabitants) and in most developed world areas, like Northern and Western Europe (6.6 and 8.4 per 100,000 male inhabitants), and Australia (8.8), with the exception of Southern Europe and Northern America (10.9 and 10.1 per 100,000 respectively).[11]

3.2.1 Etiologies and risk factors

The majority of the HCC cases occur in patients with underlying liver disease, mostly as a result of hepatitis B or C virus (HBV or HCV) infection or alcohol abuse.[12] However, several other risk factors have been associate to the development of HCC and their variable distribution is depending on geographic regions, race and ethnic group.[13] In sub-Saharan Africa and eastern Asia, around 80% of the cases develop from chronic hepatitis B (HBV) and aflatoxin B1 (AFB1) exposure. Aflatoxin, a fungal metabolite, has a strong hepatocarcinogen effect and the risk of HCC is conditioned by the dose and the time of exposure. Moreover, it has been shown that the co-effect of AFB1 and HBV virus leads to a 30 times stronger increase of HCC risk compared to AFB1 alone exposure.[14] Hepatitis C virus (HCV) and alcohol abuse, instead, are common risk factors in USA, Europe and Japan. Estimated HCC risk significantly increases in HBV and HCV infected patients and is even higher in patients with established cirrhosis.[15,16]

Even though HBV and HCV are the major causes of HCC, around 40% of patients do not show neither virus infection nor alcohol abuse, suggesting the presence of other causes of the disease. Among them, metabolic syndrome due to obesity and diabete, non-alcoholic fatty liver disease (NAFLD), hereditary hemochromatosis, $alpha_1$-antitrypsin deficiency, autoimmune hepatitis, some porphyrias, and Wilson's disease.[13] Cigarette smoking is also considered a risk factor associated with HCC development.

3.2.2 Staging system

A good staging system is important to assess the most appropriate therapeutic approach for each patient. Nevertheless, the heterogeneity and the complexity of HCC makes the classification more complicated. Indeed, nowadays, different systems are used based on several parameters like size of the tumor, lymph-node infiltration, vascular invasion, presence of metastasis and some other liver function measures.

3.2.2.1 Barcelona Clinic Liver Cancer (BCLC)

Currently, Barcelona Clinic Liver Cancer (BCLC) staging classification is the most widely recognized clinical algorithm used for HCC patient stratification and treatment allocation. It combines tumor stage and liver function parameters.[17,18,19] According to this system, patients can be classified in four stages: early stage (BCLC 0 or A), intermediate stage (BCLC B), advanced stage (BCLC C) and end stage (BCLC D) (Figure 2).

Early stage HCC (BCLC 0 or A) includes patients suitable for surgical resection and liver transplantation. It is further divided in four subgroups: stage A1 (single tumors and absence of relevant portal hypertension and normal level of bilirubin); stage A2 (single tumors associated with relevant portal hypertension and normal level of bilirubin); stage A3 (single tumors with both relevant portal hypertension and increased bilirubin); stage A4 (three tumors smaller than 3 cm independently of their liver function).

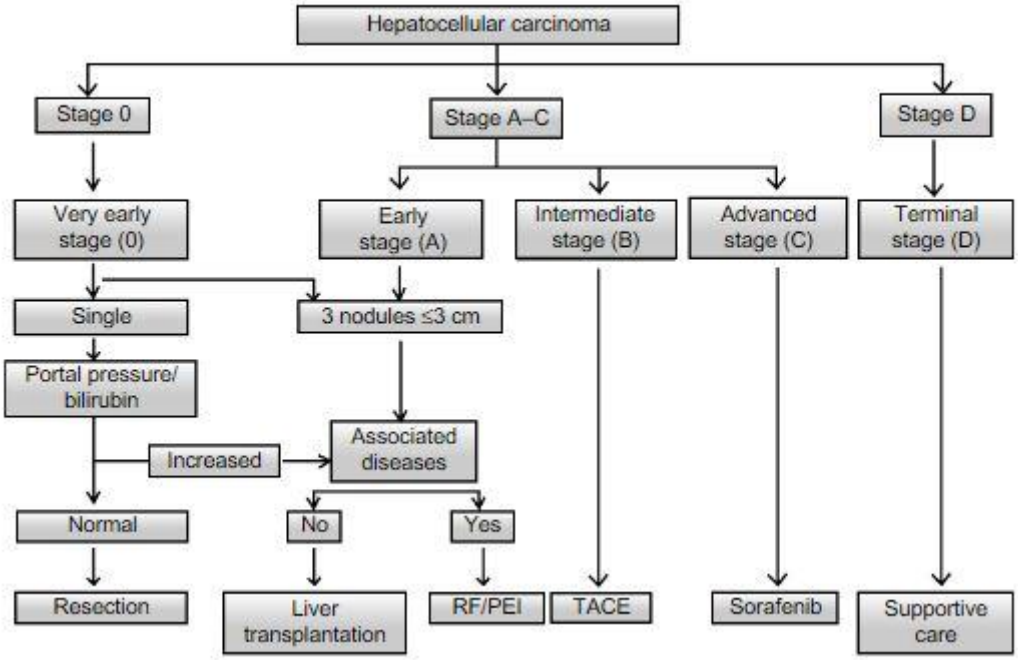


Figure 2: **Barcelona Clinic Liver Cancer (BCLC) staging classification and therapeutic strategies.** Taken from Singal et al.[20] Staging and therapeutic strategies of tumors according to the BCLD

Intermediate stage HCC (BCLC B) includes asymptomatic patients having multinodular tumors, preserved liver function and absence of vascular invasion or extrahepatic spread. Patients in this stage could benefit from transarterial chemoembolization (TACE), a non-invasive technique that allows local delivery of beads in order to reduce tumor blood supply together with local administration of chemotherapy.

Advanced stage HCC (BCLC C) consists of patients with either symptomatic tumors or with an invasive tumoral scenario characterized by vascular invasion or extrahepatic spread. Patients in this group are the best candidates to receive new antitumoral agents including the tyrosine kinase inhibitor Sorafenib.

End stage HCC (BCLC D) includes patients with severe symptoms and irreversible liver functional impairment where only palliative treatment can be administered.

The major advantage of the BCLC system is that it can be used to identify patients with early-stage HCC, who may benefit from curative therapies, contrary to patients with advanced-stage disease who would benefit more from palliative treatment.

3.2.2.2 Other Staging systems

Many other staging systems may be employed and used in the HCC classification. The Okuda system[21] uses parameters related to the liver functional status (albumin, ascites, bilirubin) and to the tumor size (> or < 50%). Using this classification, patients can be grouped in three stages: stage I (including patients with better prognosis compared to the other two stages), stage II and stage III. The negative remark of the Okuda classification is that it is mainly used for patients with advanced-end tumor stage and it cannot discriminate between early and advanced stage.

The Child-Pugh system gives a score from 1 to 3 (3 is the most severe) to five clinical measures : total bilirubin, serum albumin, prothrombin time, ascites and hepatic encephalopathy.[22] Based on the total score, patients are then classified in three groups: A, B or C. Group A, the less severe, includes a score of 5-6 points and 5-year survival rate of 95%; group B is the moderate one with a score of 7-9 points and a 5-year survival rate of 75%; finally, group C with 10-15 points and 50% of 5-year survival rate is the most severe disease.[23]

The Cancer of the Liver Italian Program (CLIP)[24] score combines four different variables: the Child-Pugh stage (A,B or C), tumor morphology (uni or multinodular with <50% extension), presence of portal vein thrombosis and serum level of alpha-fetoprotein (AFP < or > 400 ng/ml). These parameters are then classified in six categories, with numbers that range from 0 (patients with good prognosis) to 6.

Relevant to mention is the Tumor-Node-Metastasis (TNM) classification.[25] In contrast with the other HCC classification systems, TNM uses only tumor characteristics without taking into account the functional capacity of the liver. Indeed, this staging system evaluates primary tumor features (T), the presence/absence of lymph nodes (N) and the distant metastasis (M).[26] Even though TNM is often still used by surgeons for assessing the success of surgical resection and liver transplantation, it has been questioned and almost abandoned for lack of prognostic value.

3.2.3 Prevention-Diagnosis-Treatments

Vaccination against HBV infection is the first line for prevention of HCC. The World Health Organization recommends the vaccination to all the newborns and to the high risk subjects in all the countries.[27] Generally, anti-HBV vaccine is administered in three doses, the second dose is given after one month from the first dose and the third one is given after six months from the second.[27] Several studies have found evidences of efficacy of HBV vaccine and reduce incidence of HCC.[28] Similarly, treatment of HCV leads to a decreased on HCC incidence.[29] Nowadays, interferon therapy is known to be used to reduce the risk of HCC in patients carrying HCV infection. However, the risk cannot be completely removed in patients with severe fibrosis or cirrhosis.[30] New direct-acting antiviral agents (DAAs) have been generated and adopted. They are antiviral drugs that interfere with HCV replication processes

like protein assembly or polymerase activity. DAAs have higher rates of response in patients with minimal adverse effects.[31] However, recent studies have been published describing the relationship between DAAs treatment and HCC recurrence. Indeed, *Reig et al* [32] and *Conti et al* [33] analyzed two different patients cohorts and both have found high HCC recurrence despite a good rate of sustained virologic response (SVR) after DAAs treatment. The results of these studies have risen the question about the risk and benefit of DAAs.

Change in lifestyle could become also another relevant line of prevention. In particular, patients having alcohol disorders can, at least, prevent the incidence of alcohol-associated cirrhosis by adopting abstinence behavior.

Defining the tumor stage of a patient is essential to address the right therapeutic procedure. Among the adopted therapies, remarkable are the surgical resection, liver transplantation, percutaneous ethanol injection (PEI), radiofrequency ablation (RFA), trans-arterial chemoembolization (TACE), radioembolization and Sorafenib treatment.

Surgical resection is used for patients in early HCC (BCLC 0 or A) having single nodules, good liver condition and either no cirrhosis or well-compensated cirrhosis. This technique aims to completely extirpate the tumor, preserving a sufficient hepatic function. Unfortunately, around 70% of patients undergoing to surgical resection develop recurrence and they need to receive additional treatment. Generally, early tumor recurrence, within two years from surgery, is mainly due to local invasion and intrahepatic metastasis. On the contrary, late recurrence happens because of *de novo* tumor formation.[34]

Liver transplantation (LT) is considered the best treatment of choice for HCC. Eligible patients for LT are those having one lesion < 5 cm or up to 3 lesions (tumors) of 3 cm each or smaller in diameter (*Milan criteria*).[35] The overall survival rate after 5-years is around 75% and, compared to liver resection, the recurrence is less than 15%.

Percutaneous ethanol injection (PEI) is an ablation technique that employs alcohol (ethanol) to kill the liver tumor cells. In particular, ethanol spreads into tumor cells thanks to a needle guided by ultrasound or computed tomography. Once diffused, the alcohol is able to cause dehydration and protein denaturation, leading to coagulative necrosis, microvascular thrombosis and final tumor ischemia. It is a safe, cheap and easy technique but it needs more than one session. PEI is used for treatment of patients with early-intermediate HCC stage (BCLC 0-B).

Radiofrequency ablation (RFA) induces thermal injury and damage of the cancer tissue using a radiofrequency energy. Indeed, the current reaches the target tissue generating heat at the site and this leads to irreversible cellular changes like protein denaturation and necrosis of tumor cells.[36]

Trans-arterial chemoembolization (TACE) aims to reduce or completely block the blood supply to the tumor by delivering specific beads. Moreover, a chemotherapy drug is also released at the target site. TACE can be performed if the liver function is good and if there are no evidences of problem with the portal veins. Patients with intermediate-stage HCC (BCLC B) can benefit from TACE.

Radioembolization follow the same principle of the TACE and can be employed in BCLC B patients. Indeed, embolization is used to block the blood flow within the tumor but this time, the procedure is in combination with radiotherapy (beads coated with yttrium-90).

Patients in advanced HCC stage (BCLC C) are eligible only for systemic treatments, like Sorafenib (Nexavar, Bayer) administration. Sorafenib is a multi-tyrosine kinase inhibitor given as first line HCC treatment. It is able to suppress tumor cell proliferation and angiogenesis and promotes tumor cell apoptosis. A phase III randomized trial, the SHARP trial, showed an improved survival from 7.9 months to 10.7 in patients treated with Sorafenib compared to the placebo group.[37] Unfortunately, due to the genetic heterogeneity of HCC, few patients can really benefit from this therapy and some of them develop resistance to Sorafenib. Elucidating possible mechanisms that lead to Sorafenib resistance can help to find strategies able to prevent or overcome the resistance, when it occurs. *EGFR* activation[38], *c-Jun* activation[39], *AKT* activation[40], hypoxic environment[41], EMT[42] (epithelial mesenchymal transition) are some of the mechanisms responsible for Sorafenib resistance. Recently, progress has been done in HCC drug development. Indeed, new drugs were approved or are likely to be approved for first and second-line treatment. Regorafenib (Stivarga, Bayer), for example, was approved for second line treatment; Nivolumab (Opdivo, Bristol-Myers Squibb) had the approval in the United States; Lenvatinib (Lenvima, Eisai) was approved in the United States and in Japan for first line treatment and it will be approved soon in Europe; Cabozantinib (Cabometyx, Exelixis) can have the approval soon for second and third line treatment.[43]

3.2.4 Genomic landscape of HCC

Hepatocarcinogenesis is a complex cascade of multistep events and histologic transformation of normal hepatocytes to HCC.[44] Typically, the tumorigenesis process starts from cirrhotic hyperplastic nodules of regenerating hepatocytes that can progress and develop into pre-malignant dysplastic lesions. Those lesions have altered cytological characteristics like clear cell changes, increase number of nuclei (nuclear crowding) as well as altered liver architecture. At this step, the dysplastic nodules can give rise to HCC.[44] From a molecular point of view, different somatic genetic and epigenetic mutations occur, followed by deregulation of several key molecular targets (Figure 3).[45] In this scenario, next generation sequencing techniques have played a fundamental role to identify the key driver mutated genes and mechanisms promoting the development of HCC from cirrhosis.[46] Exome sequencing revealed that, on average, HCCs harbor 30-40 non-synonymous mutations in the exome, but few are expected to be driver mutations.[47,48]

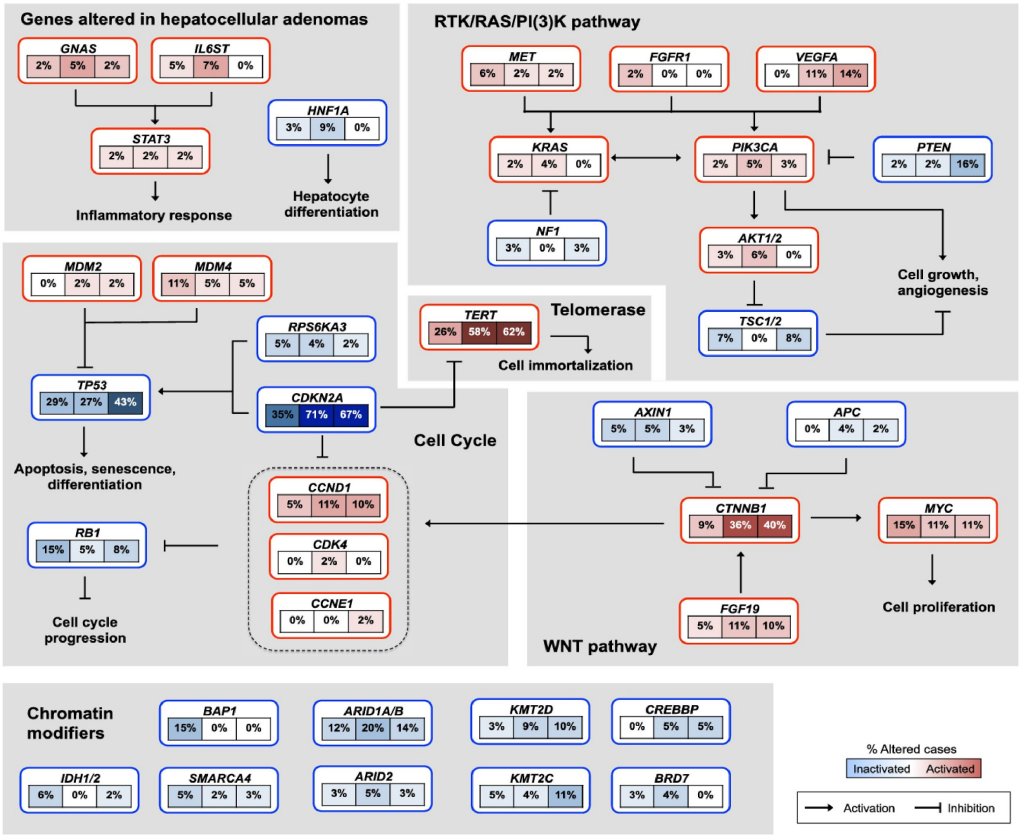


Figure 3: Representation of the most common mutated genes and pathways in HCC. Taken and modified from The Cancer Genome Atlas's study.[49]
Frequency of the somatic alterations in HCC, grouped by their involvement in different molecular pathways.

The most recurrent mutated genes in HCC include *TP53* (p53), *PIK3CA*, and *CTNNB1* (*ß-catenin)*[50] mutated in 15-40% of patients. Several biological processes and pathways are also mutated, including chromatin remodeling (*ARID1A*, *ARID1B*, *ARID2*, *BAP1*, *MLL*, *MLL3*, *PBRM1*), Wnt/β-catenin pathway (e.g. *CTNNB1*, *AXIN1*) and response to oxidative stress (e.g. *KEAP1*, *NFE2L2*).[48,51] Moreover, the landscape of mutations changes among HCCs with different etiologies. Indeed, HBV-associated HCCs have mutations in the Wnt/β-catenin and JAK/STAT pathways in 65.2% and 45.5% of cases respectively; 62.5% of HCV-associated HCCs harbor *CTNNB1* mutations[48] and alcohol-associated HCCs show mutations on chromatin remodeling genes.[47,48] Interestingly, 44-59% of HCCs have mutations in

the *Telomerase Reverse Transcriptase* (*TERT*) promoter and these mutations are associated with *CTNNB1* mutations[52] suggesting a cooperation of the two pathways in the hepatocarcinogenesis. It has been also shown that *TERT* promoter mutations are the most recurrent genetic alterations in pre-neoplastic lesions, found in 6% of low-grade dysplastic nodules, 19% of high-grade dysplastic nodules, reaching 61% of early HCCs.[53] To define the molecular landscape of HCC, it's is important to mention also the copy number alterations (CNAs) with frequent gains of 1q, 5, 6p, 7, 8q, 17q, and 20, and losses of 1p, 4q, 6q, 8p, 13q, 16, 17p and 21.[47] Although few amplifications were identified like in *CCND1* and *FGF19*, recurrent homozygous deletions affecting known tumor suppressor genes such as *CDKN2A*-*CDKN2B* (6.4%), recurrently mutated gene *AXIN1* (3.2%) and novel cancer gene *IRF2* (3.2%), have been found.[47]

3.2.5 Molecular subtypes of HCC

In the last decade, a significant proportion of HCCs samples have been analyzed and characterised on molecular and transcriptomic level, thus allowing the identification of new sub-groups of tumors.[54,55] In 2013, Hoshida et al. performed a meta-analysis of a wide number of gene expression profiles data coming from 9 independent HCC cohorts (8 previously studied and 1 reported in the mentioned study), with a total of 603 patients.[54] Three HCC subgroups, named S1,S2 and S3, resulted from the analysis and each of them correlated with specific clinical and molecular features. From a clinical point of view, S1 was histologically similar to S2 with higher risk of HCC early recurrence and an invasive fingerprint; S2 subgroup was characterized by larger tumors compared to the other groups having increased level of the serum HCC biomarker alpha fetoprotein (AFP); S3 included smallest and

the majority of well differentiated tumors. Therefore, S1 and S2 were associated with poor survival; on the contrary, S3 included tumors with better clinical outcome.[54] From a molecular perspective, instead, S1 was strongly associated with activation of WNT and TGFβ pathway as well as EMT related genes; S2 signature was due to *MYC* and *AKT* activation as well as downregulation of IFN-related genes; finally, S3 showed activation of p53 and p21 target genes and enrichment of *CTNNB1* mutation.[54] Prior to this subclassification, Boyault et al. characterized, on a transcriptomic level, 57 HCCs, 3 hepatocellular adenomas and 5 pools of non tumoral tissues.[56] In this work, they classified tumor samples in six subgroups, from G1 to G6. The group G1-G3 was mainly associated with chromosomal instability contrary to G4-G6 that were chromosome stable. In detail, G1 grouped tumors having low HBV copies and carrying overexpression of genes involved in development and parental imprinting; G2 included tumors with high HBV copies, mutations in *PIK3CA* and *TP53;* G3 tumors overexpressed genes involved in cell-cycle regulation and, as G2 group, showed *TP53* mutations without HBV infection; G4 represented the heterogeneous group and include also the 5 pools of non tumoral samples; G5 and G6 carried *CTNNB1* mutations and activation of WNT pathway and G6 group alone was found associated with satellite nodules and downregulation of cell adhesion proteins.[56] Recently, an updated molecular classification was proposed by the Cancer Genome Atlas Research Network.[49] They characterized 363 HCCs based on somatic mutations and copy number; in addition to this large cohort, other 196 patients were added and analyzed their DNA methylation profile, microRNA and protein expression. Three molecular subgroups were coming out from the analysis: iClust1, 2 and 3. iClust 1 included tumors with high vascular invasion and tumor grade, low frequency of *CDKN2A silencing,* low

percentage of *CTNNB1* and *TERT* mutation; iClust2 and 3, on the contrary, showed *TERT* mutation, *CTNNB1* and *HNF1A* mutation. Compared to iClust1, iClust2 is characterized by low microvascular invasion and tumor grade. Finally, iClust3 is associated with chromosomal instability with frequent loss of 17p, high percentage of tumors having *TP53* mutation and hypomethylation of CpGs.[49]

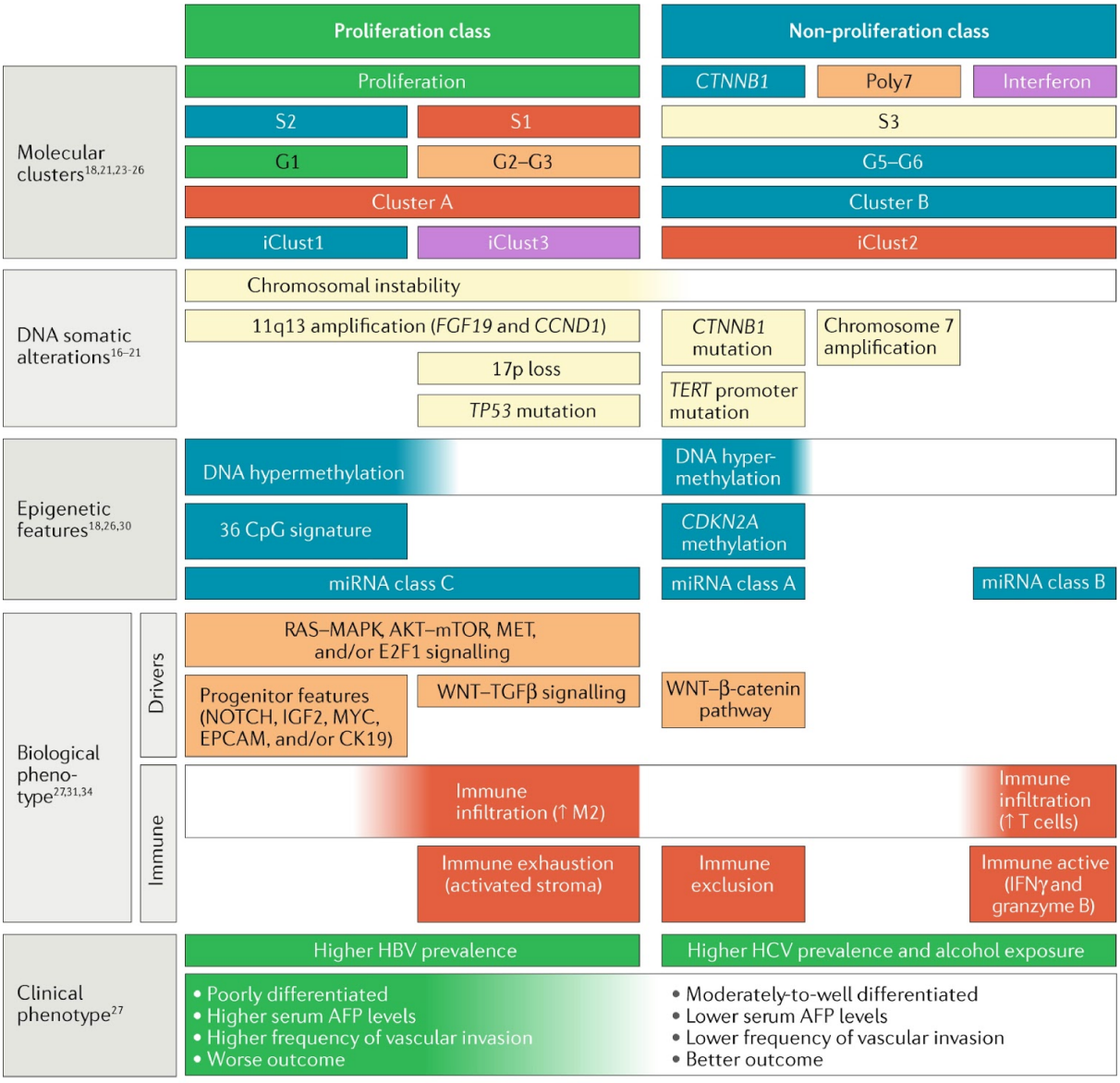


Figure 4: Summary of the different molecular HCC classifications. Taken and modified from Llovet et al.[57]
Proliferative class (left) and non-proliferative class (right) including all the different HCC molecular classifications based on molecular clusters, DNA somatic alterations, epigenetic features, biological and clinical phenotype.

3.3 TEAD proteins

TEADs proteins are a group of transcription factors (TFs) originally discovered through a genetic mosaic screen in *Drosophila* and further studied for their important role in organ development.[58] While *Drosophila* has only one *TEAD* gene, *Scalloped*, in mammals there are four highly conserved members of the family (*TEAD1, TEAD2, TEAD3, TEAD4*).

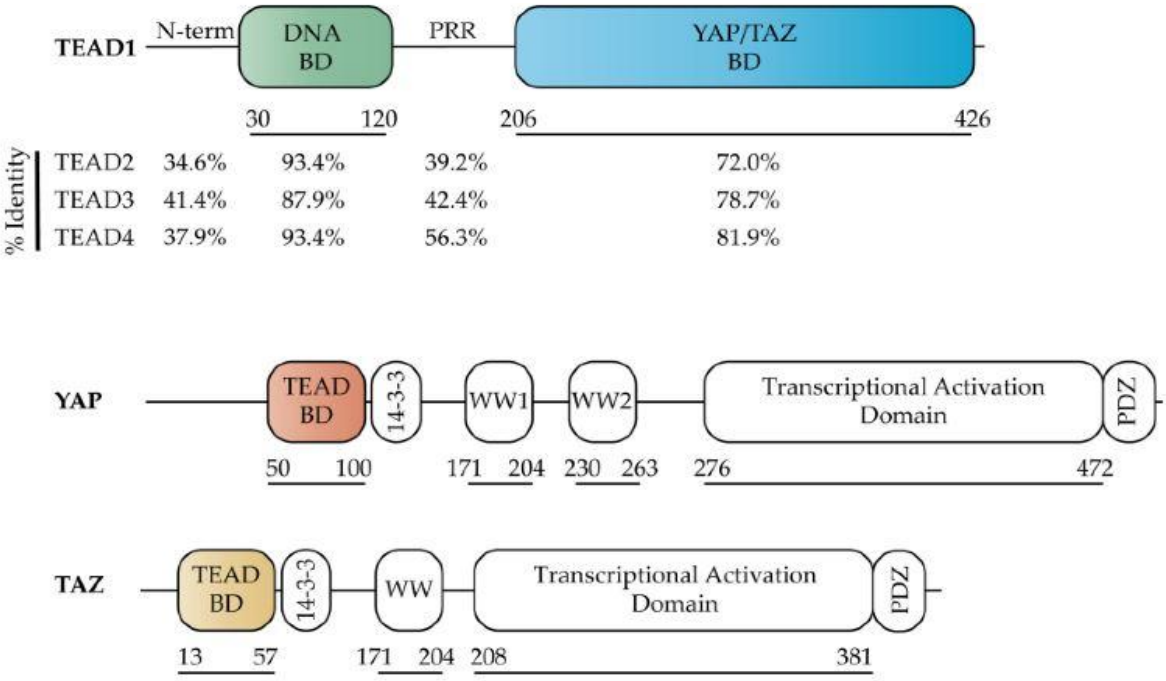


Figure 5: Structure of TEADs genes and their co-activators YAP/TAZ. Taken from Holden et al.[59] Percentage of homology among the members of TEAD family (upper panel) and schematic representation of the structure of TEAD1, YAP and TAZ (lower panel).

Structurally, all of them share a common DNA binding domain (DBD) and YAP binding domain (YBD) (Figure 5).[60] The DNA binding domain (TEA/ATTS) is located at the N-terminus while the YBD is at C-terminus and is remarkable for the binding of specific coactivators. Indeed, it has been proved that TEADs family alone is not able to activate the gene expression of their target genes but all the members require additional coactivators to accomplish their function.[61] Therefore, the coactivators do

not bind directly the DNA but, thanks to their activation domain, they help the interaction of the TF with the transcriptional machinery (Fig 5). Different TEADs coactivators have been discovered, including YAP, TAZ, the vgll proteins and the p160 family of nuclear receptor coactivators. Despite the high percentage of homology among the family, TEADs proteins are differently expressed in the tissues and in different phases of the development process. Almost all the tissues express at least one *TEAD* gene, while a few others have high levels of all four of them. Functional studies in mice shed light on the different functions of the family. TEAD1 has an impact in heart biogenesis and in the differentiation of cardiac muscle,[62] TEAD2 is important at the embryonic stage,[63] TEAD3 is more specific for the placenta and finally TEAD4 is mostly expressed in the skeletal muscle. Moreover, it has been shown that *TEAD1* and *TEAD2* double mutant embryos have even more adverse aberrations than the single *TEAD1* or *TEAD2* single mutant, showing severe growth defects, morphological alterations and dying at embryonic day 9.5.[64] Additionally, knockdown of *TEAD4* has also a fatal role during the embryos development by blocking selectively the formation of trophectoderm cell lineage.[65] *TEAD3* deficient mice, on the other hand, have not been described yet thus remaining unclear the exact function of TEAD3 during development.

3.3.1 TEADs and Hippo Pathway

Despite the relevant function of the TEADs family during development, our knowledge of their role has extensively developed in the contest of the Hippo signaling pathway. The Hippo Pathway was originally discovered in *Drosophila Melanogaster* and then it was widely studied in mammals. The core of this pathway is a kinase cascade that starts from the phosphorylation and activation of MST1/2 by TAO kinases 1/2/3.[66] MST1/2, phosphorylates and activates two scaffold proteins

SAV1 (Salvador Homolog 1) and MOB1A/B that help MST1/2 in the activation of LATS1/2 (large tumor suppressor 1/2).[67] The activation of LATS1/2 is followed by the phosphorylation of YAP (Yes-associated protein) and TAZ (transcriptional coactivator with PDZ-binding motif) proteins that are degraded and made unable to translocate to the nucleus to bind the TEADs family (Figure 6, left). Indeed, when the Hippo pathway is inactive, the entire phosphorylation process do not occur; thus the YAP and TAZ proteins remain unphosphorylated and free to migrate in the nucleus to bind TEADs and regulate the transcription of the downstream target genes (Figure 6, right).[68] In the nucleus, TEADs proteins can also bind to VGLL4 (vestigial like family member 4), acting as a repressor of the transcription. When YAP and TAZ are then active, they compete with VGLL4 for the binding to TEADs, dissociating VGLL4 and activating the transcription of the target genes. Specifically, many of the genes involved in different processes such as proliferation, migration, differentiation, epithelial-to-mesenchymal transition (EMT), apoptosis and invasion represent the main group of genes targeted by the Hippo Pathway and the TEADs proteins.[69]

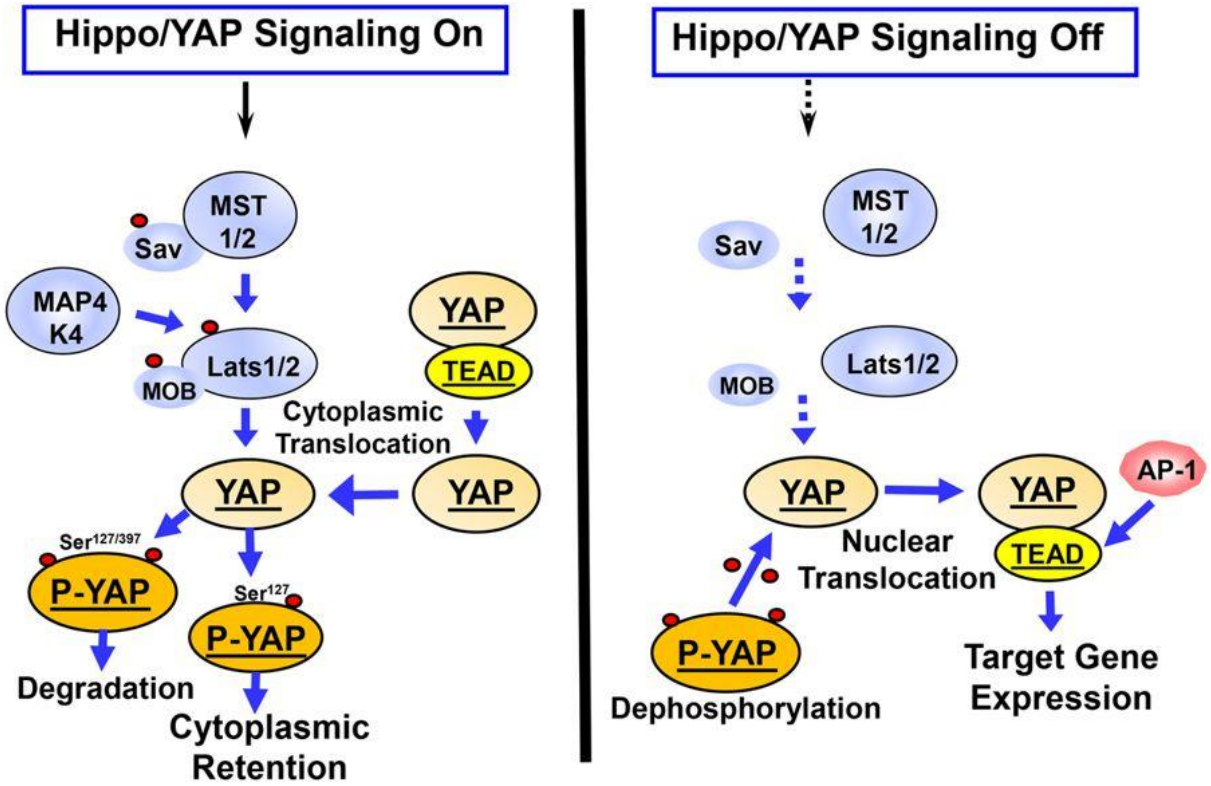


Figure 6: Hippo signaling cascade. Taken from Rozengurt et al.[70]
Schematic representation of the Hippo signaling cascade. When the Hippo pathway is ON (Figure 6, left) in response to stimuli such as cell density, the phosphorylation starts from MST1/2 and reaches YAP. Phosphorylation of YAP promotes its cytoplasmic retention or its degradation. When the Hippo pathway is OFF (Figure 6, right), YAP is dephosphorylated and can migrate to the nucleus, bind TEAD transcription factors and stimulate the transcription of the respective target genes.

3.3.2 TEADs and cancer

Dysregulation of *TEAD* genes correlate with tumorigenesis. Particularly, overexpression of TEADs was found in different types of tumors including breast, lung, prostate, colon, melanoma, glioblastoma and gastric cancer.[71] For example, in colorectal cancer, it has been reported the impact of TEAD1 in promoting cells proliferation.[72] An additional work on colon cancer, demonstrated that high *TEAD4* expression promotes epithelial-mesenchymal transition (EMT) and migration *in vitro* and metastasis formation *in vivo*.[73] Moreover, the oncogenic potential of TEAD4 was

also elucidated in breast cancer, in particular in triple negative breast cancer subtype.[74] In this specific group, indeed, it has been demonstrated that TEAD4 in cooperation with the oncogenic transcription factor KLF5, promotes breast cancer cells proliferation through the inhibition of the cyclin-dependent kinase inhibitor (CDKi) *p27* transcription activity.[74] CDKi are known to regulate cell cycle arrest, acting as tumor suppressor in some tumor types.[75]

The upregulation of TEAD factors has been shown to regulate the expression of pro-growth factors, such as *CTGF* (connective tissue growth factor),[76] Cyr61, receptor tyrosine kinase AXL,[77] Myc and survivin. For example, several TEAD-binding motifs have been found on the promoter of *CTGF,* two on the promoter region of *Cyr61*[78] and four on the transcription starting site region of *AXL.*[77]

In prostate cancer, gastric cancer and colorectal cancer, TEADs levels is used as a prognostic marker due to their positive correlation with poor clinical outcome.[73,79] On the contrary, there are also few types of tumor like bladder, renal cancer and some breast cancers, that show decreased TEADs expression, as described in the ONCOMINE database.[80–82]

Due to the recognized role of TEAD factors in tumorigenesis, they have been proposed as potential therapeutic targets.[83,59] Unfortunately, the challenge is still ongoing because all the possible inhibitor candidates should be able to reach the nucleus to strongly and specifically bind TEAD. In addition to that, it is not yet well elucidated whether the need is to develop a panel of inhibitors specific for every member of the family or one inhibitor could be sufficient, due to the high homology

among TEAD1-4 proteins.[59] Pobbati et al. shed light on the possibility to target the YAP-binding domain (YBD) of TEADs, having a peculiar hydrophobic and druggable pocket.[84] Among different FDA-approved drugs, they identified Flufenamic acid (FA) as a good compound with the best affinity for TEAD pocket, blocking then TEAD-YAP dependent regulation of processes like proliferation and migration.[84] So far, inhibitors of the Hippo Pathway have been approved: Dasatinib and Pazopanib, for example, target YAP and TAZ by promoting their phosphorylation and proteasomal degradation; C19 instead induces the activation of the MST1 and LATS1 kinases blocking the Hippo Pathway; Statins block YAP and TAZ in the cytoplasm.[85]

3.3.3 TEAD4

Located on chromosome 12, *TEAD4* is one of the four members of the TEF family. It is preferentially expressed in the skeletal muscle[86] but its expression was also found in the heart, placenta and some other organs. It binds M-CAT regulatory elements on the promoter of muscle specific genes to regulate their expression and transcription.[8788] As for all the other members of the family, TEAD4 plays a key role during development. In particular, studies in mice have showed its expression at very early stage (eight-cells) of development and embryons having no functional *TEAD4 (TEAD4 -/-)* stopped their development at morula stage, without forming the blastocoel.[89] TEAD4 is considered the main downstream effector of the previous described Hippo Pathway, essential in organ size control, stem cell maintenance and tissue homeostasis [90]. Moreover, due to the lack of a transactivation domain, TEAD4 requires co-activator factors, like the well known members of the Hippo Pathway cascade YAP and TAZ, to regulate the transcription of the downstream target genes.

In 2016 it was described a splicing isoform of TEAD4, TEAD4-S that lacks the N-terminal DNA binding domain, but maintain the YAP interaction domain, thus leading to attenuation of YAP-TEAD4 signaling.[91] The discovery of this alternative splicing mechanism provided a new possible approach for cancer therapies because promoting its re-expression could improve survival and decrease cancer cell proliferation and tumor growth. Indeed, *TEAD4* was reported to act as an oncogene in cancer, being found mostly upregulated in several tumor types like breast and gastric cancer,[71] thus promoting *TEAD4* as a good therapeutic and prognostic candidate.

4. AIM OF THE THESIS

Our group at the Institute of Pathology combines computational biology tools and molecular biology techniques to discover and elucidate mechanisms driving cancer initiation and progression. We believe that a multi-modality approach incorporating multi-omics with molecular biology enhances advances not only to define clinically relevant predictive biomarkers of response to therapy, but also to discover novel drug targets for patients unlikely to respond to standard therapies.

This approach is made possible by the availability of tissue specimens in our valuable tissue biobank that allows us to validate our in vitro findings in samples derived from patients. My project had the great opportunity to be developed taking advantage of the above mentioned expertise and resources.

The first part of my PhD thesis focused on the expression analysis of *TEAD4* in hepatocellular cancer, using publicly available dataset on one hand, and tissue specimens on the other hand. These included re-analysis of The Cancer Genome Atlas dataset[49] and the chromatin immunoprecipitation performed by the Encode project[92] on HCC cell lines. Using the datasets we could find the transcriptome and the interactome of *TEAD4* in HCC.

The second part of my PhD was the *in vitro* characterization of *TEAD4* expression in hepatocellular carcinoma cell lines in terms of oncogenicity using molecular biology techniques to define proliferation and migration of cells overexpressing *TEAD4*.

Last but not least we identified and validated a new TEAD4 target involved in tumorigenesis that may be explored as novel therapeutic target.

5. MATERIALS AND METHODS

5.1 Immunohistochemistry

Immunohistochemical staining of TEAD4 was assessed on a tissue microarray consisting of a cohort of 434 patient specimens from resection, as previously described.[93,94] After excluding samples for which the tissue punch was absent or had poor staining quality, 192 HCCs, 108 cirrhotic tissues and 79 normal liver samples were available.[93] IHC was performed using anti-TEAD4 antibody (Abcam; clone ab97460; dilution 1:100). Staining was performed on a Benchmark immunohistochemistry staining system (Ventana Medical Systems, Inc., Tucson, AZ) using iVIEW-DAB as chromogen. Immunoreactivity was scored semi-quantitatively as the number of positive tumor cells (nuclear staining) over the total number of tumor cells as well as the intensity of the signal by an experienced hepatopathologist (Prof. Luigi M Terracciano).[93]

5.2 Cell lines

HCC-derived cell lines (HLE and HUH7) were maintained in a 5% CO2-humidified atmosphere at 37°C and cultured in DMEM supplemented with 10% FBS, 1% Pen/Strep (Bio-Concept) and 1% MEM-NEAA (MEM non-essential amino acids, ThermoFisher Scientific). Both cell lines were confirmed negative for mycoplasma infection using the PCR-based Universal Mycoplasma Detection kit (American Type Culture Collection, Manassas, VA) as previously described.[95]

For experiments involving HPS70 inhibitor (Heat Shock Protein Inhibitor I, 373260, Calbiochem), cells were incubated with culture media containing 100 μM of inhibitor and the corresponding DMSO control.

5.3 Plasmids and Transfection

For *TEAD4* overexpression, pLV[Exp]-EGFP/Neo-EF1A>hTEAD4 was designed and ordered on the Vector builder platform and the empty control vector was pCMV-mir EGFP. For *TEAD4* silencing, pSuper-retro-puro empty vector and pSuper-retro-puro shTEAD1/3/4 were adapted from Bin Zhao et al.[76] kindly provided by our collaborator Dr. Fengyuan Tang (Prof. Dr. Gerhard Christofori's laboratory, Department of Biomedicine, Basel). The expression vectors were transiently transfected using the jetPRIME transfection reagent (Polyplus) following the manufacturer's instructions. The expression of the plasmids was evaluated by western blot and qRT-PCR analysis. Cells were harvested 48h after transfection for further experiments.

5.4 RNA extraction and qRT-PCR

RNA from HCC cell lines and 43 frozen tissues was extracted following the Trizol® method (Invitrogen Life Technologies®), according to the manufacturer's specifications. One microgram of RNA was retro-transcribed using the SuperScript VILO cDNA synthesis kit (Thermo Fisher Scientific). Quantitative real-time PCR for the expression levels of *TEAD4* and *HSPA6* was performed with Sybr Green method. *GAPDH* was used as housekeeping gene. The fold changes in gene expression were calculated using the standard ΔΔCt method.[96] To quantitate *TEAD4* transcript levels, dilutions of *TEAD4* plasmid were used as standard curves (dilutions ranged from 10^8 to 10^0 copies of plasmid).

The primer sequences' information is listed below:

Gene	Forward primer (5'-3')	Reverse primer (3'-5')
TEAD4	GCTCCTTCTATGGGGTCTCC	GTGCTTGAGCTTGTGGATGA
HSPA6	CGTGCCCGCCTATTTCAATG	AAAAATGAGCACGTTGCGCT
GAPDH	AGGTGAAGGTCGGAGTCAACG	TGGAAGATGGTGATGGGATTT

5.5 Protein extraction and Western Blot

Proteins were extracted using Co-IP buffer (100 mmol/L NaCl, 50 mmol/L Tris pH 7.5, 1 mmol/L EDTA, 0.1% Triton X-100) supplemented with 1x protease inhibitors (cOmplete Mini, EDTA-free Protease Inhibitor Cocktail, Roche, CO, #4693159001) and 1x phosphatase inhibitors (PhosSTOP #4906837001, Merck). Cell lysates were then treated with 1x reducing agent (NuPAGE Sample Reducing Agent, Invitrogen, #NP0009), 1x loading buffer (NuPAGE LDS Sample Buffer, Invitrogen, #NP0007), boiled and loaded into neutral pH, pre-cast, discontinuous SDS-PAGE mini-gel system (NuPAGE 10% Bis-Tris Protein Gels, ThermoFisher,). The proteins were then transferred to nitrocellulose membranes using Trans-Blot Turbo Transfer System (Bio-Rad). The membranes were blocked for 1 hr with Sure Block (Lubio science) and then probed with primary antibodies overnight at 4°. Next day, the membranes were incubated for 1 hr at RT with fluorescent secondary goat anti-mouse (IRDye 680) or anti-rabbit (IRDye 800) antibodies (both from LI-COR Biosciences). Blots were scanned using the Odyssey Infrared Imaging System (LI-COR Biosciences) and band intensity was quantified using ImageJ software. The ratio of proteins of interest/loading control in treated samples were normalized to their counterparts in control cells.

5.6 Proliferation Assay

Proliferation assays were performed using the xCELLigence Real-Time Cell Analysis (RTCA, ACEA Biosciences, San Diego, CA, USA) system. Cells were first seeded and transfected in 6 well plates and 24 h after transfection $5x10^3$ cells were resuspended in 100 µl of medium and plated in each well of an E-plate 16. Background impedance was measured after adding 50 µl of the corresponding medium to each well of the E-16 plate. The final volume in each well was then 150 µl. The impedance signals were recorded every 15 minutes until 96/120 h and expressed as cell index values, calculated automatically and normalized by the RTCA Software Package v1.2. The values were defined as mean ± standard deviation. Mann-Whitney test was used for statistical analysis with GraphPad software.

5.7 Migration Assay

Migration assays were performed using the CIM-plate of the xCELLigence Real-Time Cell Analysis (RTCA, ACEA Biosciences, San Diego, CA, USA) system. Cells were first transfected in tissue flasks and 24 h after transfection, they were harvested and seeded in the CIM-plate. Every well of the bottom chamber was filled with 160 ul of the corresponding medium at 10% FBS concentration. After placing the upper chamber on top of the lower chamber, 50 µl of serum free medium was added on each CIM well for the background measurement. After 3x PBS washing, $3x10^4$ cells re-suspended in 100 µl of the corresponding medium at 1% FBS concentration were seeded in each well of the upper chamber. The measurements were taken every 15

minutes until 24/48 h after seeding and expressed as cell index values. Mann-Whitney test was used for statistical analysis with GraphPad software.

5.8 CellTiter Glo- cell viability assay

CellTiter-Glo® (G7573, Promega) was used to determine the number of viable cells based on ATP content. Cells were first seeded and transfected in 6 well plates and 24 h after transfection $5x10^3$ cells were resuspended in 100 µl of medium and plated in each well of a 96 well plates at different time points. Cell viability was measured by adding 100 µl of CellTiter-Glo® /well. After 10 min of incubation at RT, the luminescence signal is measured.

5.9 Antibodies

Antibodies are listed below:

Antibody	Company	Catalog number	Application	Dilution/Concentration
Anti-TEAD4	Abcam	ab97460	Western blot/IHC	1:500/1:100
TEF-3(TEAD4)	Santa Cruz Biotechnology	sc-101184	Chip-qPCR	10 µg
HSP70	Cell signaling	4872T	Western blot	1:500
Normal mouse IgG	Santa Cruz Biotechnology	sc-2025	Chip-qPCR	10 µg

5.10 RNA-sequencing

Ion Ampliseq™ Transcriptome Human Gene expression panel, Chef-Ready kit from Ion Torrent (Thermo Fisher Scientific) was used for library preparation following the datasheet guidelines. After RNA extraction from HUH7 cells transiently overexpressing *TEAD4* (n=4) or empty vector (n=4), the RNA samples were treated with Turbo Dnase (AM 1907, Thermo Fisher Scientific). 10 ng of RNA for each sample was then reverse transcribed using the SuperScript VILO cDNA synthesis kit (Thermo Fisher Scientific) and resuspended in a final volume of 15 µl. The cDNA was amplified for 12 cycles using the Ion AmpliSeq Transcriptome Human Gene Expression kit panel that targets over 20,000 genes. The resulting pool of libraries was then quantified by qPCR using the Ion Universal Quantification kit (Thermo Fisher Scientific). Expected dilution was around 100 pM. The pool was then diluted to a 50 pM final concentration and loaded on an Ion 540™ chip using the Ion Chef™ instrument and sequenced on an Ion S5™ instrument (Thermo Fisher Scientific). Raw data was processed directly on the Ion Torrent Server™ and aligned to the reference hg19 AmpliSeq Transcriptome fasta reference. Absolute reads matrix was downloaded from the *Ion Torrent* server and the analysis was performed using *edgeR* package.[97] Genes with low expression (< 1 log-counts per million in ≥ 2 samples) were filtered out. Normalization was performed using the “TMM” (weighted trimmed mean) method and differential expression was assessed using the quasi-likelihood F-test. Genes with false discovery rate (FDR) < 0.05 were considered differentially expressed. Gene set enrichment analysis was performed using *fgsea* package[98] with genes ranked based on logFC and p-value. Hallmark gene sets from MSigDB[99] were used to identify significantly upregulated/downregulated pathways.

5.11 Chromatin Immunoprecipitation (ChIP)

The protocol used for ChIP was previously described[100] and adapted for the experiment. Cells from four 10 cm Petri dishes at 70/80% confluence were crosslinked in 1% formaldehyde for 10 min with continuous shaking. The crosslinking was stopped by adding 0.15 M glycine while continuing shaking. After collecting the cells by scraping, the pellet was washed 3x with cold PBS. Nuclei were isolated and lysed in order to proceed with the sonication, using the Bioraptor instrument. The number of cycles and the settings were as described previously.[101] At the same time, the antibody was coupled with magnetic protein G beads (Invitrogen 100-03D) by incubating 75 µl of protein G beads with 10 µg of TEAD4 antibody (TEF-3 sc101184 Santa Cruz Biotechnology) and 10 µg of mouse IgG (sc-2025 Santa Cruz Biotechnology) as a negative control for 1h at RT with a constant rotation. At the end of the sonication process, an aliquot of the chromatin was kept as input control for every sample and an equal amount of sonicated chromatin was incubated with magnetic beads-antibody coupled previously at 4°C overnight while rotating. The samples were washed several times with different buffers and then eluted with an elution buffer (all the buffers corresponded to the ones described in the original protocol from Blecher-Gonen et al.).[100] RNase treatment first and then Proteinase K treatment were done on all the samples including the input followed by overnight reverse cross-linking at 65°C with continuous shaking. DNA purification followed using Agencourt AMPure XP (A63880 Beckman Coulter). TEAD4 abundance on specific target genes promoters was quantified by qRT-PCR compared to the IgG negative control.

Gene	Forward primer (5'-3')	Reverse primer (3'-5')
HSPA6	GGCCATTCACTAAGGAACCA	TCAGGAAGGCCGAAGATATG
CTGF	GCCAATGAGCTGAATGGAGT	CAATCCGGTGTGAGTTGATG

5.12 Analysis of The Cancer Genome Atlas (TCGA) data

mRNA-seq RSEM normalized gene level expression data for TCGA hepatocellular carcinoma cohort (n=371) were obtained from FIREBROWSE website.[102] Using *TEAD4* expression, samples with z-score > +1.96 were considered *TEAD4* upregulated while z-score < -1.96 were *TEAD4* downregulated. Pathological information used in this study were retrieved from a previous study.[103] Associations between clinicopathological parameters and *TEAD4* expression were assessed by $\chi 2$ test. For survival analysis, FPKM expression data of *TEAD4* and overall survival information were obtained from Human Protein Atlas (Pathology Atlas) liver cancer project.[104] Survival cut-off was determined using *maxstat* package[105] and analysis was performed using Kaplan-Meier method and log-rank test. Univariate and Multivariate Cox-regression analysis was performed to investigate the association between overall survival and clinicopathologic variables. All tests were two-sided and $P < 0.05$ were considered as statistically significant.

5.13 Analysis of ChiP-seq data

ChIP-seq data (accession number GSM1010875)[92] provided by the HudsonAlpha Institute for Biotechnology was obtained from Gene Expression Omnibus. ChIP was performed using mouse monoclonal immunoglobulin G, raised against recombinant

protein TEAD4, on the HepG2 human liver cancer cell line. The data, in the form of bigWig file, were loaded into Integrative Genomics Viewer[106] and the “Find Motif” tool was used to search for the most conserved bases of TEAD4 binding motif within regions of TEAD4 ChIP-seq peaks. TEAD4 binding motif was obtained from the JASPAR 2018 database.[107]

6. RESULTS

6.1 TEAD4 expression is upregulated in HCC at transcript and protein level

A total of 28% *TEAD4* genomic alterations were seen in tumors relative to normal samples in TCGA data (cutoff based on z-score greater than +1.96 are *TEAD4* upregulation and less than -1.96 are *TEAD4* downregulation) (Fig 7A). We then evaluated *TEAD4* expression in tumors compared to normal samples and found *TEAD4* over-expressed in tumors compared to normal (Fig 7B). Furthermore, survival analysis on the TCGA data revealed worse overall survival in patients with high *TEAD4* expression (p=0.002) (Fig 7C). In addition, both univariate and multivariate analyses showed that *TEAD4* expression is an independent predictor of overall survival in liver patients ($p< 0.05$) (Table1). Correlation of *TEAD4* expression with clinicopathological parameters showed that *TEAD4* overexpression is associated with poorly differentiated HCCs (Edmondson grades III and IV, p=0.002), presence of necrotic areas (p=1.92E-05), absence of Mallory bodies (p=0.004) and with cytological variants including pleomorphic cells, clear cells, multinucleate cells and fatty change (all $p<0.01$) (Table 2). No significant correlation was found with gender, race, risk factors, cholestasis, vessel infiltration, infiltrating lymphocytes and histological growth pattern. We confirmed that *TEAD4* was overexpressed in an independent cohort of 24 HCCs frozen tissues compared to 19 cirrhosis (Fig 7D). To further evaluate TEAD4 expression on the protein level, we stained and scored a tissue microarray (TMA) containing 192 HCCs, 108 cirrhotic tissues and 79 normal liver samples. The results were consistent with the ones obtained from TCGA dataset: TEAD4 positivity as assessed by the percentage of TEAD4+ cells was found in a good portion of HCCs compared to non tumoral tissues (Fig 7E), as

shown from the two representative pictures from the TMA (Fig 7F). Taken together, these data demonstrate higher TEAD4 expression in HCCs at the mRNA and protein levels.

		Overall Survival		
		Univariate Analysis	Multivariate Analysis	
Clinicopathologic Variables	**Category**	**p-value**	**p-value**	**HR (95% CI)**
Sex	Female vs Male	0,24		
TEAD4 Expression	High vs Low	**0,035**	**0,0469**	1.7890 (1.0079 - 3.175)
Vessel Infiltration		0,16		
Edmondson Grade	II vs III + IV	0,91		
Stages		**0,057**		
Infiltrating lymphocytes		0,12		

Table 1: Univariate and multivariate analysis for the effect of *TEAD4* expression on overall survival.

Clinicopathological features		LowTEAD4 expression	HighTEAD4 expression	P-value
		n (%)	n (%)	
Gender (n=358)* Low TEAD4 (n=304), High TEAD4 (n=54)	Female	95 (31.3%)	21 (38.8%)	0,26
	Male	209 (68.8%)	33 (61.1%)	
Race (n=349) Low TEAD4 (n=295), High TEAD4 (n=54)	American indian or Alaska native	1 (0.3%)	0 (0.0%)	0,77
	Asian	127 (43.1%)	27 (50.0%)	
	Black or African American	15 (5.1%)	2 (3.7%)	
	White	152 (51.5%)	25 (46.3%)	
Risk factor (n=341) Low TEAD4 (n=290), High TEAD4 (n=51)	At least one risk factor	74 (25.5%)	15 (29.4%)	0,55
	No risk factor	216 (74.5%)	36 (70.5%)	
	Alcohol consumption	95 (32.8%)	19 (37.3%)	0,53
	Absence of Alcohol consumption	195 (67.2%)	32 (62.7%)	
	Hepatitis B virus	88 (30.3%)	14 (27.5%)	0,67
	Absence of Hepatitis B virus	202 (69.7%)	37 (72.5%)	
	Hepatitis C virus	49 (16.9%)	5 (9.8%)	0,20
	Absence of Hepatitis C virus	241 (83.1%)	46 (90.2%)	
	Hemochromatosis	4 (1.4%)	1 (2.0%)	0,75
	Absence of Hemochromatosis	286 (98.6%)	50 (98.0%)	
	Non-Alcoholic Fatty Liver Disease	20 (6.9%)	0 (0.0%)	0,05
	Absence of Non-Alcoholic Fatty Liver Disease	270 (93.1%)	51 (100.0%)	

Child pugh classification grade (n=235) Low Tead4 (n=206), High Tead4 (n=29)	A	186 (90.2%)	27 (93%)	0,85
	B	19 (9.2%)	2 (6.9%)	
	C	1 (0.4%)	0 (0.0%)	
Edmondson Grade (n=359) Low Tead4 (n=305), High Tead4 (n=54)	II	116 (38.0%)	9 (16.7%)	**<0.01**
	III + IV	189 (62.0%)	45 (83.3%)	
Cholestasis (n=358)* Low TEAD4 (n=304), High TEAD4 (n=54)	Absent	234 (77.0%)	44 (81.5%)	0,46
	Present	70 (23.0%)	10 (18.5%)	
Mallory Bodies (n=359) Low TEAD4 (n=305), High TEAD4 (n=54)	Absent	229 (75.1%)	50 (92.6%)	**<0.01**
	Present	76 (24.9%)	4 (7.4%)	
Vessel infiltration (n=358)* Low TEAD4 (n=304), High TEAD4 (n=54)	Absent	202 (66.4%)	32 (59.3%)	0,30
	Present	102 (33.6%)	22 (40.7%)	
Necrotic areas (n=358)* Low TEAD4 (n=305), High TEAD4 (n=53)	Absent	244 (80.0%)	28 (52.8%)	**<0.01**
	Present	61 (20.0%)	25 (47.2%)	
Infiltrating lymphocytes (359) Low TEAD4 (n=305), High TEAD4 (n=54)	Lymphocytes	145 (47.5%)	23 (42.6%)	0,50
	Absence of lymphocytes	160 (52.5%)	31 (57.4%)	
Histological Growth Pattern (n=359) Low TEAD4 (n=305), High TEAD4 (n=54)	Trabecular	296 (97.0%)	52 (96.3%)	0,76
	Absence of Trabecular	9 (3.0%)	2 (3.7%)	

	Pseudoglandular	165 (54.1%)	30 (55.5%)	0,84
	Absence of Pseudoglandular	140 (45.9%)	24 (44.4%)	
	Compact	139 (45.6%)	29 (53.7%)	0,26
	Absence of Compact	166 (54.4%)	25 (46.3%)	
Cytological Variants (n=291) Low TEAD4 (n=244), High TEAD4 (n=47)	Pleomorphic cells	130 (53.3%)	38 (80.9%)	**<0.01**
	Absence of Pleomorphic cells	114 (46.7%)	9 (19.1%)	
	Clear cells	65 (26.6%)	6 (12.8%)	**0,04**
	Absence of Clear cells	179 (73.7%)	41 (87.2%)	
	Multinucleated cells	71 (29.1%)	24 (51.1%)	**<0.01**
	Absence of Multinucleated cells	173 (70.9%)	23 (48.9%)	
	Fatty Change	105 (43.0%)	7 (14.9%)	**<0.01**
	Absence of Fatty Change	139 (57.0%)	40 (85.1%)	
	Hyaline bodies	49 (20.1%)	9 (19.1%)	0,88
	Absence of Hyaline bodies	195 (79.9%)	38 (80.9%)	
	Pale bodies	10 (4.1%)	2 (4.3%)	0,96
	Absence of Pale bodies	234 (95.9%)	45 (95.7%)	
	Ground glass	9 (3.7%)	0 (0.0%)	0,18
	Absence of Ground glass	235 (96.3%)	47 (100.0%)	

* One case was not evaluable

P-value **< 0.05** is significant and is calculated using $\chi 2$ test

Table 2: Association between *TEAD4* expression and clinicopathological features.

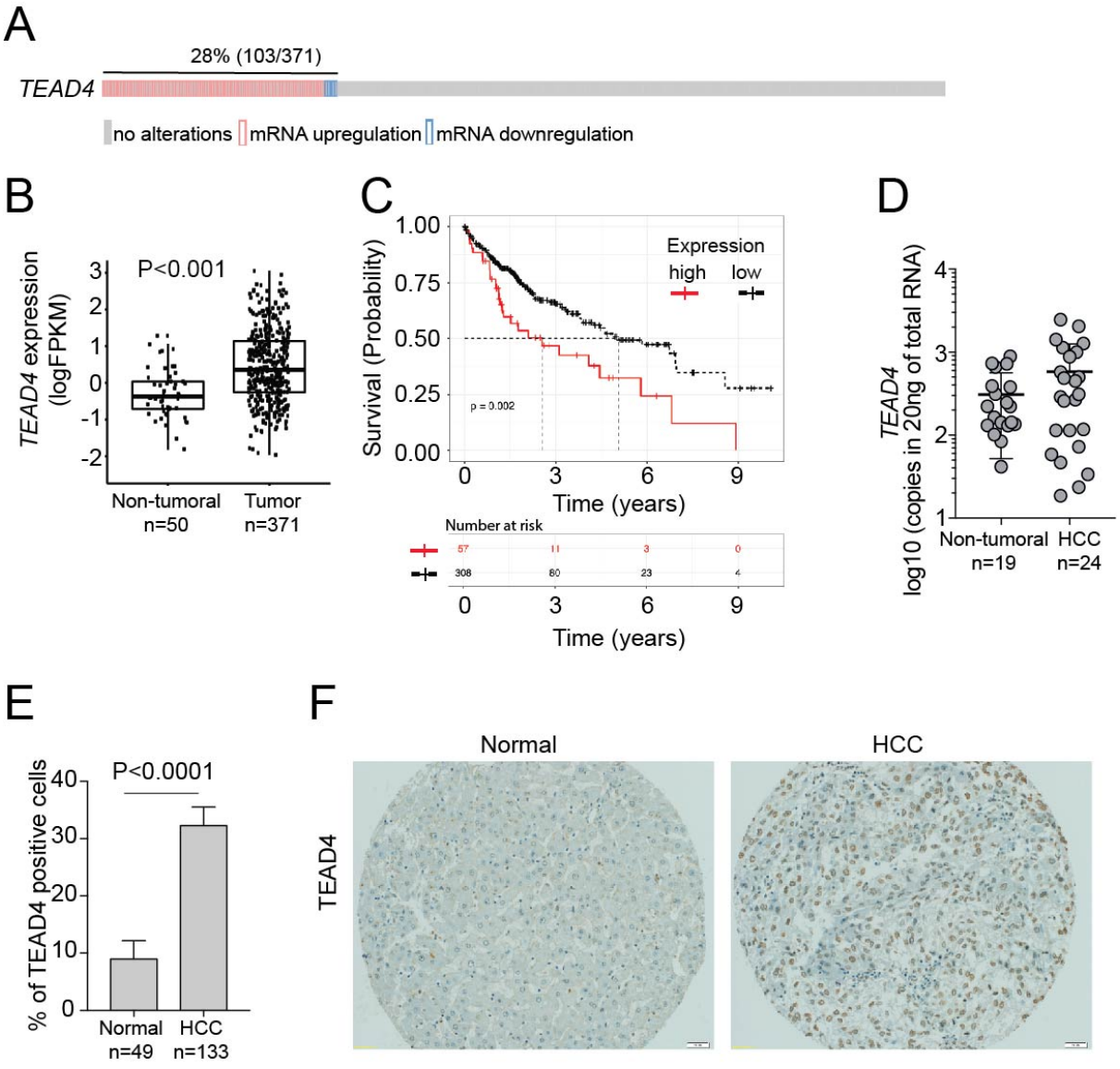
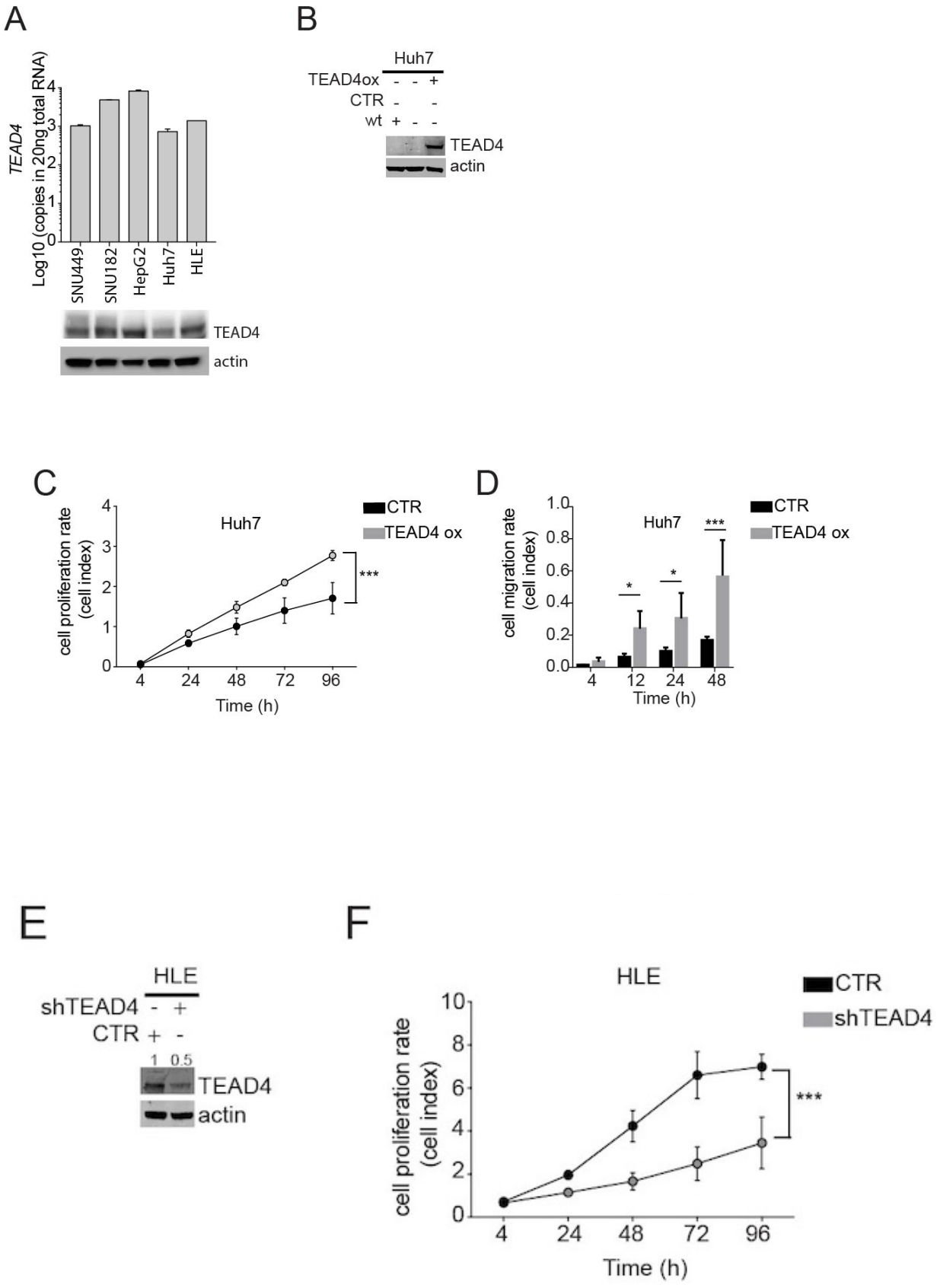


Figure 7: *TEAD4* is overexpressed in HCCs. A) Percentage of *TEAD4* dysregulation in HCC samples from TCGA data; **B)** *TEAD4* expression level in tumors (n=371) compared to non tumoral tissues (n=50) in TCGA dataset; **C)** Overall survival analysis (Kaplan-Meier) of HCC patients from TCGA dataset after stratification for *TEAD4* high and *TEAD4* low expression, p=0.002; **D)** *TEAD4* mRNA was measured in an independent cohort of 24 frozen HCC tissues compared to 19 cirrhosis. *GAPDH* was used as internal control and the results are shown as copy numbers per 20 ng of total RNA; **E)** Percentage of TEAD4 positive cells in HCC and normal tissues by IHC; **F)** Representative IHC pictures for TEAD4 staining of a normal (left) and HCC tissue (right).

6.2 TEAD4 overexpression increases tumor growth and migration *in vitro.*

To investigate the functional role of TEAD4 in hepatocarcinogenesis, we performed several *in vitro* experiments. First, we screened a panel of HCC cell lines for TEAD4 RNA and protein expression levels and we observed substantial variability among cell lines (Fig 8A). Based on these results, we selected HUH7 and HLE for *TEAD4* overexpression and silencing to study the effects of TEAD4 on cell proliferation and migration. Transfection efficiency on both cell lines was assessed by western blot analysis (Fig 8B and E). We first evaluated whether TEAD4 could have an impact on the cell proliferation rate. Using the Xcelligence system, we observed that high *TEAD4* levels led to increased proliferation capacity of overexpressing HUH7 cells (Fig 8C). On the contrary, when *TEAD4* was transiently silenced in the HLE cell line, the growth ratio was significantly reduced (Fig 8F). Similarly, we assessed the migration capacity of the cells 48h after transfection using transwell Xcelligence plates. Forced *TEAD4* expression in HUH7 resulted in increased cell migration capacity (Fig 8D) and on the other hand, *TEAD4* knockdown reduced migration activity in HLE (Fig 8G). Taken together, these data confirmed the oncogenic role of TEAD4 as reported in other cancer types, showing its importance in enhancing cell growth and migration in liver cancer cells.


A
TEAD4
Log10 (copies in 20ng total RNA)
SNU449
SNU182
HepG2
Huh7
HLE
TEAD4
actin
B
Huh7
TEAD4ox - - +
CTR - -
wt + - -
TEAD4
actin
C
Huh7
CTR
TEAD4 ox
cell proliferation rate (cell index)
***
Time (h)
D
Huh7
CTR
TEAD4 ox
cell migration rate (cell index)
*
*
***
Time (h)
E
HLE
shTEAD4 - +
CTR + -
1 0.5
TEAD4
actin
F
HLE
CTR
shTEAD4
cell proliferation rate (cell index)
***
Time (h)

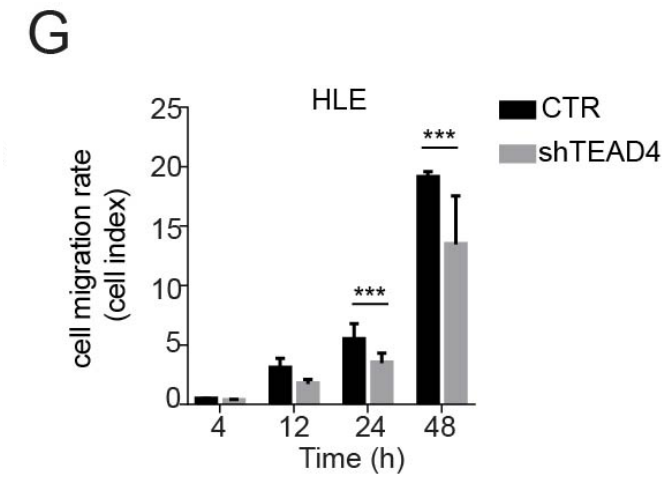


Figure 8: TEAD4 enhances proliferation and migration *in vitro.* A) Western blot analysis of TEAD4 overexpression in HUH7. β-actin was used as loading control; **B-C)** HUH7 cell line was transiently transfected with a plasmid overexpressing *TEAD4* and the corresponding CTR (control) vector. Proliferation **(B)** and migration **(C)** were assessed by the xCELLigence system at the times indicated; **D)** Western blot confirming TEAD4 silencing in HLE cell line. β-actin was used as loading control; **E-F)** HLE cell line was transiently transfected with shTEAD4 vector and the corresponding shCTR. Proliferation **(E)** and migration **(F)** were assessed by the xCELLigence system at the times indicated.

6.3 RNA-sequencing shed light on a new TEAD4 target gene *HSPA6*

The oncogenic functions of TEAD4 in liver cancer cells led us to discover and characterize potential novel TEAD4 targets involved in cancer progression. We therefore performed RNA sequencing on HUH7 cells transiently overexpressing *TEAD4* (n=4) or empty vector (n=4). Transfection efficiency of two biological and two technical replicates was evaluated by western blot (Fig 8A).

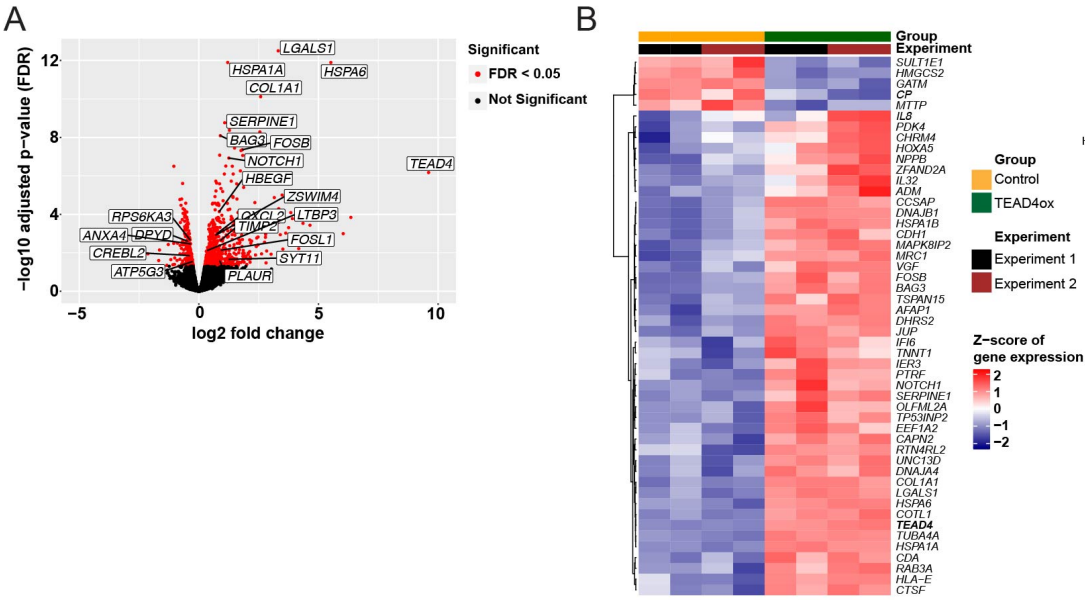
Differential gene expression analysis was performed to identify genes that were upregulated and downregulated upon *TEAD4* overexpression, using false discovery rate < 0.05 as the threshold for statistical significance. As shown in the volcano plot (Fig 9A), 569 genes were upregulated and 316 were downregulated. The heatmap

showing the expression pattern of the top 50 differentially expressed genes demonstrated the homogeneity and robustness between the replicates (Fig 9B).
As shown in Fig 9C, gene set enrichment analysis was performed to identify the most significantly enriched pathways upon *TEAD4* overexpression. Interestingly, we observed that hallmark pathways such as myogenesis, P53 pathway and EMT were significantly upregulated (FDR < 0.05) (Fig 9D). By contrast, pathways involved in protein secretion, bile acid metabolism were downregulated (padj < 0.05). These results are consistent with the more aggressive phenotype we observed upon *TEAD4* overexpression.

Interestingly, one of the most significantly upregulated gene was *HSPA6 (*Heat Shock Protein Family A (Hsp70) Member 6), a member of the ubiquitous and highly conserved Hsp70 family of molecular chaperones. Several studies have demonstrated the oncogenic role of the Hsp70 family and their overexpression was found in different types of cancer.[108,109,110,111] Additionally, the correlation of HSP70 expression with higher tumor grade and poor prognosis[112,113] as well as with increased cell proliferation rate and malignancy[114–116] have also been reported.

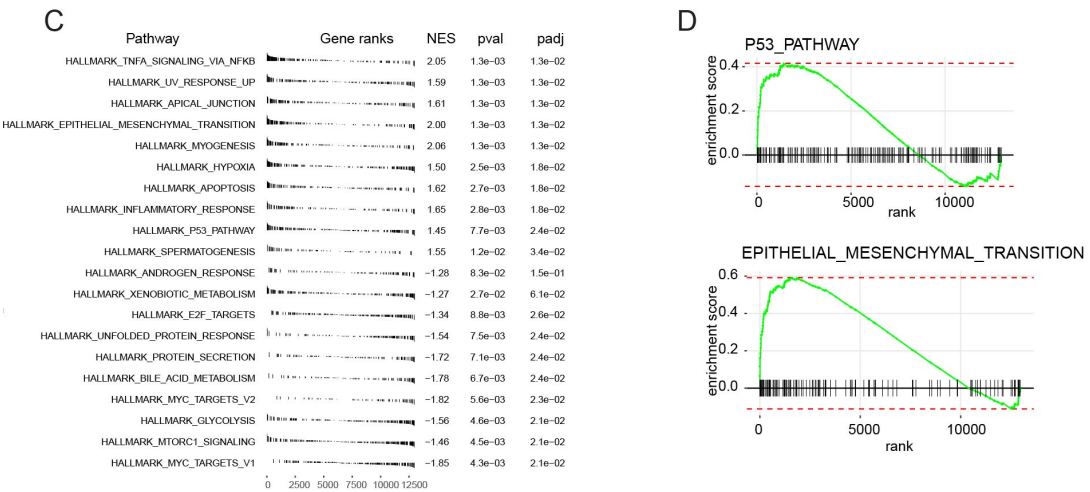
To ensure that the changes in gene expression was specifically due to TEAD4 modulation, RNA sequencing was also performed on the HLE cell line transiently silenced with shTEAD4 and the corresponding shCTR. Transfection efficiency was previously evaluated by western blot (Fig 8D). As shown in the volcano plot (Fig 9E), 108 genes were upregulated and 179 were downregulated; *HSPA6* was found to be downregulated upon *TEAD4* silencing, consistent with the results of *TEAD4* overexpression. Heatmap in Fig 9F showed the top 50 differentially expressed

genes. Pathway analysis performed on these samples (Fig 9G) showed that the pathways enriched upon *TEAD4* silencing showed the opposite trend to those enriched upon *TEAD4* overexpression. Specifically, myogenesis, P53 pathway and EMT were downregulated upon *TEAD4* silencing (Fig 9H). In conclusion, our results are consistent with the phenotypic characterization of TEAD4 and suggested *HSPA6* as a new TEAD4 target gene.



C

Pathway	NES	pval	padj
HALLMARK_TNFA_SIGNALING_VIA_NFKB	2.05	1.3e−03	1.3e−02
HALLMARK_UV_RESPONSE_UP	1.59	1.3e−03	1.3e−02
HALLMARK_APICAL_JUNCTION	1.61	1.3e−03	1.3e−02
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	2.00	1.3e−03	1.3e−02
HALLMARK_MYOGENESIS	2.06	1.3e−03	1.3e−02
HALLMARK_HYPOXIA	1.50	2.5e−03	1.8e−02
HALLMARK_APOPTOSIS	1.62	2.7e−03	1.8e−02
HALLMARK_INFLAMMATORY_RESPONSE	1.65	2.8e−03	1.8e−02
HALLMARK_P53_PATHWAY	1.45	7.7e−03	2.4e−02
HALLMARK_SPERMATOGENESIS	1.55	1.2e−02	3.4e−02
HALLMARK_ANDROGEN_RESPONSE	−1.28	8.3e−02	1.5e−01
HALLMARK_XENOBIOTIC_METABOLISM	−1.27	2.7e−02	6.1e−02
HALLMARK_E2F_TARGETS	−1.34	8.8e−03	2.6e−02
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	−1.54	7.5e−03	2.4e−02
HALLMARK_PROTEIN_SECRETION	−1.72	7.1e−03	2.4e−02
HALLMARK_BILE_ACID_METABOLISM	−1.78	6.7e−03	2.4e−02
HALLMARK_MYC_TARGETS_V2	−1.82	5.6e−03	2.3e−02
HALLMARK_GLYCOLYSIS	−1.56	4.6e−03	2.1e−02
HALLMARK_MTORC1_SIGNALING	−1.46	4.5e−03	2.1e−02
HALLMARK_MYC_TARGETS_V1	−1.85	4.3e−03	2.1e−02



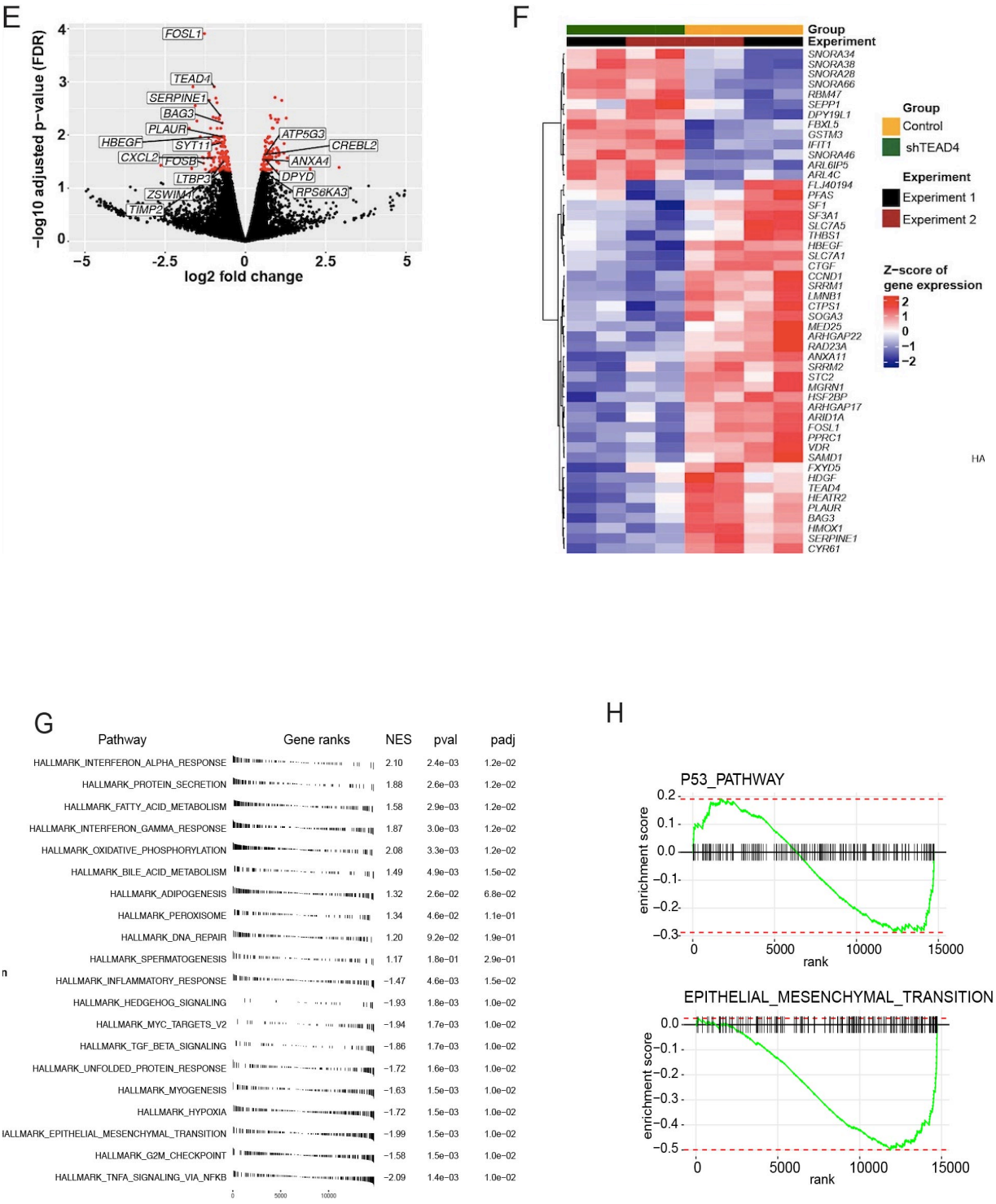


Figure 9: RNA-seq revealed *HSPA6* as a new TEAD4 target gene. A,E) Volcano plot showing genes differentially expressed in **A)** HUH7 TEAD4 ox versus CTR (FDR < 0.05) and **E)** HUH7 shTEAD4 versus CTR (FDR < 0.05); **B,F)** gene expression heatmap showing the top 50 differentially expressed genes in **B)** HUH7 TEAD4 ox versus CTR and **F)** HUH7 shTEAD4 versus CTR; **C,G)** gene set enrichment analysis in **C)** HUH7 TEAD4ox and **G)** HUH7 shTEAD4. Genes ranked based on logFC and p-value; **D,H)** hallmark gene set identified enrichment plots for P53 and EMT

pathways in **D)** HUH7 TEAD4ox and **H)** shTEAD4. y-axis represents the enrichment score.

6.4 TEAD4 regulates proliferation by binding the promoter of *HSPA6*

To determine whether *HSPA6* was a direct transcriptional target of TEAD4, we downloaded and analyzed the ChIP-seq data (GSE32465)[92] available for the HEPG2 cell line used in a genome-wide study of YAP/TAZ/TEAD4 binding sites from Zanconato F et al.[117] Interestingly, as shown in Fig 10A, a TEAD4 peak with two regions harboring the TEAD4 binding motif (chr1:161493541-161493545 and chr1:161493648-161493652) is located on the promoter region of *HSPA6*, around 100 bp upstream of the transcription starting site (TSS).

Given our finding that *TEAD4* modulation alters the expression of *HSPA6* and that TEAD4 binds to *HSPA6* promoter region,[117] we wanted to clarify the mechanism of interaction of TEAD4 and *HSPA6*. First, we wanted to confirm the direct binding of the transcription factor on the promoter region of *HSPA6* gene in our own model systems by performing chromatin immunoprecipitation (ChIP) assay. Therefore HUH7 transiently overexpressing *TEAD4* were cross-linked and immunoprecipitated with anti-TEAD4 or anti-IgG antibodies. Immunoprecipitation of chromatin was then analyzed by quantitative PCR (qPCR), using primers spanning the predicted TEAD4 binding region on the *HSPA6* promoter. As with the *CTGF* promoter, a positive control that is a known TEAD4 binding target,[86] we observed TEAD4 occupancy on the *HSPA6* promoter (Fig 10B).

We next sought to investigate the downstream effects of the direct regulation of TEAD4 on the *HSPA6* promoter. Upon *TEAD4* overexpression in HUH7, RNA was

collected at different time points and both *TEAD4* and *HSPA6* expression were measured by qRT-PCR. As shown in Fig 10C, *TEAD4* mRNA expression was the highest at 48h after transfection and its overexpression led to increased *HSPA6* expression, showing the highest level as well at 48h after transfection. We also evaluated the protein level of HSP70B (the protein product of *HSPA6*) in HUH7 overexpressing *TEAD4* by western blot and showed that HSP70B protein increased upon *TEAD4* overexpression (Fig 10D).

Given that previous studies have demonstrated that HSP70 is overexpressed and enhances proliferation, invasion and metastasis in various cancer types, including HCC,[118] we asked whether the oncogenic effects of TEAD4 is mediated through its upregulation of HSP70B. We therefore tested this hypothesis by inhibiting HSP70B in HUH7 transiently overexpressing *TEAD4*. In *TEAD4*-overexpressing HUH7 cells treated with 100 μM of Heat Shock Protein Inhibitor I 24 h after transfection, we observed a strong reduction in cell viability (Fig 10E), especially at 48h after treatment. Taken together, our results demonstrated that TEAD4 promotes carcinogenesis via the direct regulation of HSP70B.

Overall, these data shed light on a new direct regulation of TEAD4 on HSP70B promoter leading to their downstream cooperation in promoting HCC proliferation *in vitro*.

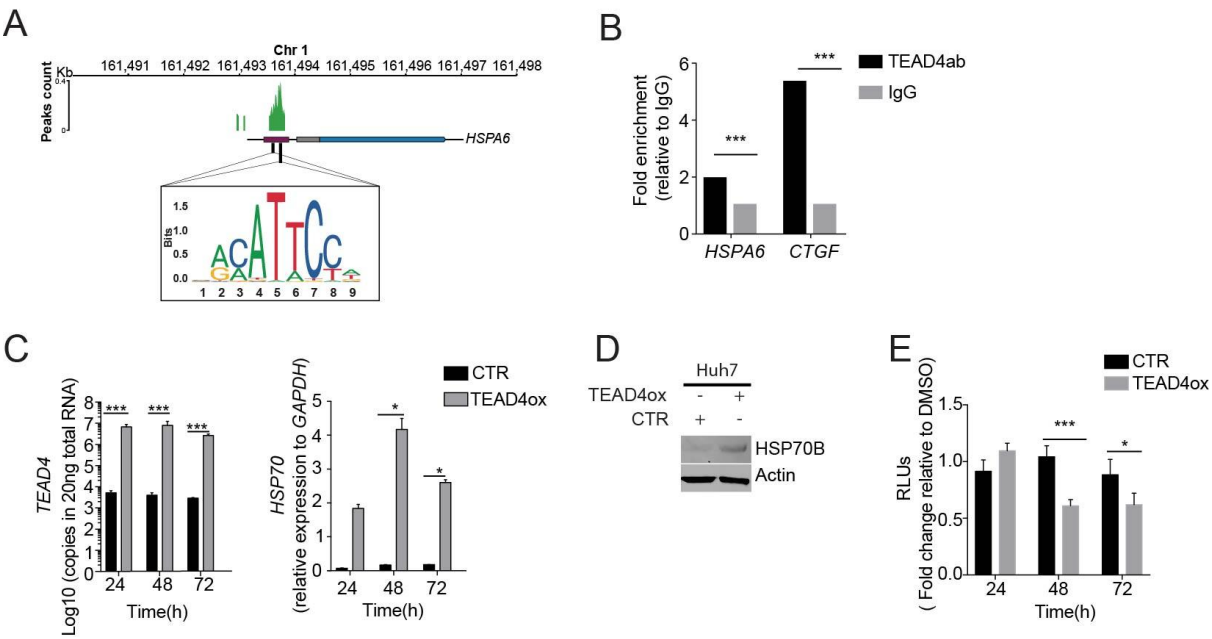


Figure 10: TEAD4 directly binds HSP70 promoter, thus impacting cell proliferation. A) Representation of TEAD4 ChIP-peak in the promoter region of *HSPA6.* ChIP-seq analysis data (accession number GSM1010875), performed on HepG2 cell line and previously analyzed from Zanconato et al.,[117] was collected from GEO (Gene Expression Omnibus). Figure of the peak was generated using pyGenomeTracks. Below, the TEAD4 sequence logo obtained from JASPAR transcription factor binding profile database is shown with the positions of the two regions harboring the motif in the *HSPA6* promoter. **B)** ChIP-qPCR showing TEAD4 enrichment on *HSPA6* promoter. *CTGF* was used as a positive control. Data from one representative experiment; **C)** HUH7 cell line was transiently transfected with a vector overexpressing *TEAD4* and the corresponding CTR. RNA was extracted at different time points. mRNA expression of *TEAD4* (left) and *HSP70 (right)* was measured by qRT-PCR using *GAPDH* as internal control. Results for *TEAD4* expression are shown as copy numbers per 20 ng of total RNA. Data derived from three independent experiments; **D)** Western blot analysis showing increased HSP70 protein level in HUH7 TEAD4ox compared to CTR, βactin was used as loading control; **E)** Viability of HUH7 TEAD4ox treated with Heat Shock Protein Inhibitor I at different time points was measured using the CellTiter-Glo® Assay. DMSO was used as control. One representative experiment is showed.

7. DISCUSSION

In the present study, we reported the effect and the impact of TEAD4 expression in liver carcinogenesis. TEAD4 is a member of the transcriptional enhancer factor family composed of four members (TEAD1-4). Deregulation of its expression, as for the other members of the family, was found in different tumor types, demonstrating its strong oncogenic function. Indeed, its role in promoting cell growth, migration and tumorigenesis by regulating the expression of target genes have been extensively reported.[74,83,119]

In the present study, we wanted to analyze the expression and the function of TEAD4 in HCC that was not well elucidated so far. From the analysis of the TCGA dataset, we found *TEAD4* upregulation on the mRNA level in 26% of HCC tumors. Given the diverse molecular profiles of HCC[57,49,120], it was not unexpected to find such a variability in *TEAD4* expression and dysregulation among tumor samples. Furthermore, *TEAD4* was upregulated in HCC tumors compared to normal samples and its expression was significantly associated with worse overall survival. In accordance with these findings, we observed a similar percentage of TEAD4 overexpression by analyzing a TMA and a small independent cohort of frozen tissues, confirming the high heterogeneity among samples. Clinicopathological correlation showed a significant association of *TEAD4* overexpression with poorly differentiated tumors (Edmonson grade III and IV), absence of necrotic areas and Mallory bodies and with cytological variants (pleomorphic cells, clear cells, multinucleate cells and fatty change). Our results are in line with other studies in human cancers, reporting the correlation of *TEAD4* expression with poor prognosis.[83,121,122]

Our next step was then to elucidate the effect of TEAD4 and its mechanism in *in vitro* models. Indeed, previous studies had shown that TEAD4 triggers oncogenic transformation in different tumor types. For example, in colorectal cancer, *TEAD4* knockdown inhibits cell proliferation *in vitro* and *in vivo.*[121] Another study in gastric cancer confirmed the key role of TEAD4 in cancer proliferation and invasiveness demonstrating that *TEAD4* and *TEAD1* silencing prevented the growth of tumor xenografts in nude mice [83]. Here we made similar observations; we demonstrated that *TEAD4* overexpression promoted tumor cell growth and migration in liver cancer cell lines. Notably, the effect was reverted when *TEAD4* was silenced.
Overall, these findings supported the known oncogenic function of TEAD4 in other tumor types, adding HCC to the group of tumors having high TEAD4 expression.

As a transcription factor, TEAD4 regulates the transcription of different genes, activating or repressing downstream pathways. For example, it is known that *CTGF* is one of its target genes and this direct regulation has an impact on cell growth.[76] Nevertheless, the landscape of targets is not entirely well known yet. Our project had indeed the goal to further elucidate the mechanism of action of TEAD4 and to discover new targets involved in HCC progression. From the analysis of gene expression profiling of HUH7 overexpressing *TEAD4* using RNA-seq, we identified 569 genes upregulated and 316 downregulated. *HSPA6 (a member of the HSP70 protein family)* was one of the most upregulated genes from the analysis. Interestingly, gene set enrichment analysis showed some important altered pathways upon *TEAD4* overexpression, including the P53 pathway, EMT and myogenesis. Moreover, we took advantage of publicly available ChIP-seq data that

were previously used by Zanconato F. et al.[117] They performed a genome wide study to characterize the allocation of YAP, TAZ and TEAD4 binding site on the human genome and they found that the vast majority of the peaks localized far from the transcription starting site (TSS) but a small fraction mapped close (around 1kb) to it. Moreover, their analysis demonstrated that this pattern is conserved in different types of cancer, like breast and liver cancers. We re-analyzed ChIP-seq data from the HepG2 cell line and interestingly we found a TEAD4 binding site mapped less than 1Kb from the TSS of *HSPA6*. We confirmed this direct binding in HUH7 by performing a ChIP-qPCR. HSP70B, the protein product of *HSPA6*, is a member of the ubiquitous and highly conserved Hsp70 family of molecular chaperones. In normal conditions, this chaperone protects cells against stress and ensures the correct folding and transport of new synthesized proteins.[123] In aberrant conditions, such as in tumors and malignant transformation, it was found dysregulated and altered. Indeed, several cancers, including breast,[124] colon,[125] liver[126] and prostate[127] cancers have reported high levels of HSP70 showing a strong positive correlation with increased proliferation rate and malignant transformation. Notably, Chuma et al, identified HSP70 as a molecular marker of early stage HCC.[128] Later on, Di Tommaso et al showed that HSP70, together with glutamine synthetase and glypican 3, is included into a panel of recognized putative markers of malignancy routinely used for the diagnosis of HCC.[129] All these findings allowed us to hypothesize that the oncogenic effect of TEAD4 could act via its regulation of HSP70B. In particular, we demonstrated that, by treating HUH7 transiently overexpressing *TEAD4* with an HSP70 inhibitor, the number of viable cells significantly decreased. Despite the fact that further insights into the molecular mechanism is needed, for instance, whether the proposed mechanism is Hippo

pathway dependent, our findings suggest a possible new targetable and druggable pathway in HCC.

In conclusion, this study identified an oncogenic role of TEAD4 in liver hepatocarcinogenesis, proposing a new mechanism of regulation on cancer proliferation by the directly regulating HSP70B.

8. REFERENCES

1. Elias, H. ANATOMY OF THE LIVER. *The Liver* 41–59 (1963). doi:10.1016/b978-1-4832-2824-2.50008-x
2. Trefts, E., Gannon, M. & Wasserman, D. H. The liver. *Curr. Biol.* **27**, R1147–R1151 (2017).
3. Burt, A. D., Ferrell, L. D. & Hubscher, S. G. *MacSween's Pathology of the Liver E-Book*. (Elsevier Health Sciences, 2017).
4. Rombouts, K. Hepatic Stellate Cell Culture Models. in *Stellate Cells in Health and Disease* 15–27 (2015).
5. Kupffer Cells in the Liver. in *Comprehensive Physiology* (ed. Terjung, R.) **20**, 453 (John Wiley & Sons, Inc., 2013).
6. Braet, F. & Wisse, E. Structural and functional aspects of liver sinusoidal endothelial cell fenestrae: a review. *Comp. Hepatol.* **1**, 1 (2002).
7. Gordillo, M., Evans, T. & Gouon-Evans, V. Orchestrating liver development. *Development* **142**, 2094–2108 (2015).
8. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **136**, E359–86 (2015).
9. Kumar, M., Zhao, X. & Wang, X. W. Molecular carcinogenesis of hepatocellular carcinoma and intrahepatic cholangiocarcinoma: one step closer to personalized medicine? *Cell Biosci.* **1**, 5 (2011).
10. Parkin, D. M., Bray, F., Ferlay, J. & Pisani, P. Global Cancer Statistics, 2002. *CA: A Cancer Journal for Clinicians* **55**, 74–108 (2005).
11. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).

12. Llovet, J. M., Burroughs, A. & Bruix, J. Hepatocellular carcinoma. *Lancet* **362**, 1907–1917 (2003).
13. El-Serag, H. B. Hepatocellular carcinoma. *N. Engl. J. Med.* **365**, 1118–1127 (2011).
14. Liu, Y. & Wu, F. Global burden of aflatoxin-induced hepatocellular carcinoma: a risk assessment. *Environ. Health Perspect.* **118**, 818–824 (2010).
15. Yang, J. D. *et al.* Cirrhosis is present in most patients with hepatitis B and hepatocellular carcinoma. *Clin. Gastroenterol. Hepatol.* **9**, 64–70 (2011).
16. El-Serag, H. B. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* **142**, 1264–1273.e1 (2012).
17. Llovet, J. M., Brú, C. & Bruix, J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. *Semin. Liver Dis.* **19**, 329–338 (1999).
18. Llovet, J. M. *et al.* Hepatocellular carcinoma. *Nat Rev Dis Primers* **2**, 16018 (2016).
19. Pons, F., Varela, M. & Llovet, J. M. Staging systems in hepatocellular carcinoma. *HPB* **7**, 35–41 (2005).
20. Singal, A. & Ravi, S. Regorafenib: an evidence-based review of its potential in patients with advanced liver cancer. *Core Evidence* 81 (2014). doi:10.2147/ce.s48626
21. Okuda, K. *et al.* Natural history of hepatocellular carcinoma and prognosis in relation to treatment. Study of 850 patients. *Cancer* **56**, 918–928 (1985).
22. Pugh, R. N., Murray-Lyon, I. M., Dawson, J. L., Pietroni, M. C. & Williams, R. Transection of the oesophagus for bleeding oesophageal varices. *Br. J. Surg.* **60**, 646–649 (1973).
23. Child-Pugh Score for Chronic Liver Disease and Cirrhosis. *Healthline* (2018).

Available at: https://www.healthline.com/health/child-pugh-classification. (Accessed: 26th April 2019)

24. Llovet, J. M. & Bruix, J. Prospective validation of the Cancer of the Liver Italian Program (CLIP) score: a new prognostic system for patients with cirrhosis and hepatocellular carcinoma. *Hepatology* **32**, 679–680 (2000).
25. Amin, M. B. *et al.* The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more 'personalized' approach to cancer staging. *CA: A Cancer Journal for Clinicians* **67**, 93–99 (2017).
26. Subramaniam, S., Kelley, R. K. & Venook, A. P. A review of hepatocellular carcinoma (HCC) staging systems. *Chin Clin Oncol* **2**, 33 (2013).
27. World Health Organization. Hepatitis B vaccines: WHO position paper, July 2017 - Recommendations. *Vaccine* **37**, 223–225 (2019).
28. Kao, J.-H. Hepatitis B vaccination and prevention of hepatocellular carcinoma. *Best Practice & Research Clinical Gastroenterology* **29**, 907–917 (2015).
29. Colombo, M. & Lleo, A. The impact of antiviral therapy on hepatocellular carcinoma epidemiology. *Hepat Oncol* **5**, HEP03 (2018).
30. Papatheodoridis, G. V., Papadimitropoulos, V. C. & Hadziyannis, S. J. Effect of interferon therapy on the development of hepatocellular carcinoma in patients with hepatitis C virus-related cirrhosis: a meta-analysis. *Aliment. Pharmacol. Ther.* **15**, 689–698 (2001).
31. Feld, J. J. & Foster, G. R. Second generation direct-acting antivirals – Do we expect major improvements? *Journal of Hepatology* **65**, S130–S142 (2016).
32. Reig, M. *et al.* Unexpected high rate of early tumor recurrence in patients with HCV-related HCC undergoing interferon-free therapy. *J. Hepatol.* **65**, 719–726 (2016).

33. Conti, F. *et al.* Early occurrence and recurrence of hepatocellular carcinoma in HCV-related cirrhosis treated with direct-acting antivirals. *J. Hepatol.* **65**, 727–733 (2016).
34. Colecchia, A. *et al.* Prognostic factors for hepatocellular carcinoma recurrence. *World J. Gastroenterol.* **20**, 5935–5950 (2014).
35. Mazzaferro, V. *et al.* Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *N. Engl. J. Med.* **334**, 693–699 (1996).
36. Ansari, D. & Andersson, R. Radiofrequency ablation or percutaneous ethanol injection for the treatment of liver tumors. *World J. Gastroenterol.* **18**, 1003–1008 (2012).
37. Llovet, J. M. *et al.* Sorafenib in advanced hepatocellular carcinoma. *N. Engl. J. Med.* **359**, 378–390 (2008).
38. Ezzoukhry, Z. *et al.* EGFR activation is a potential determinant of primary resistance of hepatocellular carcinoma cells to sorafenib. *International Journal of Cancer* **131**, 2961–2969 (2012).
39. Chen, W. *et al.* Activation of c-Jun predicts a poor response to sorafenib in hepatocellular carcinoma: Preliminary Clinical Evidence. *Scientific Reports* **6**, (2016).
40. Chen, K.-F. *et al.* Activation of Phosphatidylinositol 3-Kinase/Akt Signaling Pathway Mediates Acquired Resistance to Sorafenib in Hepatocellular Carcinoma Cells. *Journal of Pharmacology and Experimental Therapeutics* **337**, 155–161 (2011).
41. Liang, Y. *et al.* Hypoxia-mediated sorafenib resistance can be overcome by EF24 through Von Hippel-Lindau tumor suppressor-dependent HIF-1α inhibition

in hepatocellular carcinoma. *Hepatology* **57**, 1847–1857 (2013).

42. Du, B. & Shim, J. Targeting Epithelial–Mesenchymal Transition (EMT) to Overcome Drug Resistance in Cancer. *Molecules* **21**, 965 (2016).
43. Rimassa, L. Drugs in Development for Hepatocellular Carcinoma. *Gastroenterol. Hepatol.* **14**, 542–544 (2018).
44. Farazi, P. A. & DePinho, R. A. Hepatocellular carcinoma pathogenesis: from genes to environment. *Nature Reviews Cancer* **6**, 674–687 (2006).
45. Forner, A., Llovet, J. M. & Bruix, J. Hepatocellular carcinoma. *Lancet* **379**, 1245–1255 (2012).
46. Llovet, J. M., Villanueva, A., Lachenmayer, A. & Finn, R. S. Advances in targeted therapies for hepatocellular carcinoma in the genomic era. *Nat. Rev. Clin. Oncol.* **12**, 436 (2015).
47. Guichard, C. *et al.* Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat. Genet.* **44**, 694–698 (2012).
48. Cleary, S. P. *et al.* Identification of driver genes in hepatocellular carcinoma by exome sequencing. *Hepatology* **58**, 1693–1702 (2013).
49. Cancer Genome Atlas Research Network. Electronic address: wheeler@bcm.edu & Cancer Genome Atlas Research Network. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* **169**, 1327–1341.e23 (2017).
50. Villanueva, A., Newell, P., Chiang, D. Y., Friedman, S. L. & Llovet, J. M. Genomics and signaling pathways in hepatocellular carcinoma. *Semin. Liver Dis.* **27**, 55–76 (2007).
51. Fujimoto, A. *et al.* Whole-genome sequencing of liver cancers identifies

etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nature Genetics* **44**, 760–764 (2012).

52. Nault, J. C. *et al.* High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. *Nat. Commun.* **4**, 2218 (2013).
53. Nault, J. C. *et al.* Telomerase reverse transcriptase promoter mutation is an early somatic genetic alteration in the transformation of premalignant nodules in hepatocellular carcinoma on cirrhosis. *Hepatology* **60**, 1983–1992 (2014).
54. Hoshida, Y. *et al.* Integrative Transcriptome Analysis Reveals Common Molecular Subclasses of Human Hepatocellular Carcinoma. *Cancer Research* **69**, 7385–7392 (2009).
55. Lee, J.-S. *et al.* A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells. *Nature Medicine* **12**, 410–416 (2006).
56. Boyault, S. *et al.* Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* **45**, 42–52 (2007).
57. Llovet, J. M., Montal, R., Sia, D. & Finn, R. S. Molecular therapies and precision medicine for hepatocellular carcinoma. *Nat. Rev. Clin. Oncol.* **15**, 599–616 (2018).
58. Xu, T., Wang, W., Zhang, S., Stewart, R. A. & Yu, W. Identifying tumor suppressors in genetic mosaics: the Drosophila lats gene encodes a putative protein kinase. *Development* **121**, 1053–1063 (1995).
59. Holden, J. & Cunningham, C. Targeting the Hippo Pathway and Cancer through the TEAD Family of Transcription Factors. *Cancers* **10**, 81 (2018).
60. Li, Z. *et al.* Structural insights into the YAP and TEAD complex. *Genes & Development* **24**, 235–240 (2010).

61. Xiao, J. H., Davidson, I., Matthes, H., Garnier, J.-M. & Chambon, P. Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1. *Cell* **65**, 551–568 (1991).

62. Chen, Z., Friedrich, G. A. & Soriano, P. Transcriptional enhancer factor 1 disruption by a retroviral gene trap leads to heart defects and embryonic lethality in mice. *Genes Dev.* **8**, 2293–2301 (1994).

63. Kaneko, K. J., Cullinan, E. B., Latham, K. E. & DePamphilis, M. L. Transcription factor mTEAD-2 is selectively expressed at the beginning of zygotic gene expression in the mouse. *Development* **124**, 1963–1973 (1997).

64. Sawada, A. *et al.* Redundant Roles of Tead1 and Tead2 in Notochord Development and the Regulation of Cell Proliferation and Survival. *Molecular and Cellular Biology* **28**, 3177–3189 (2008).

65. Yagi, R. *et al.* Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development* **134**, 3827–3836 (2007).

66. Boggiano, J. C., Vanderzalm, P. J. & Fehon, R. G. Tao-1 phosphorylates Hippo/MST kinases to regulate the Hippo-Salvador-Warts tumor suppressor pathway. *Dev. Cell* **21**, 888–895 (2011).

67. Hergovich, A., Schmitz, D. & Hemmings, B. A. The human tumour suppressor LATS1 is activated by human MOB1 at the membrane. *Biochem. Biophys. Res. Commun.* **345**, 50–58 (2006).

68. Meng, Z., Moroishi, T. & Guan, K.-L. Mechanisms of Hippo pathway regulation. *Genes & Development* **30**, 1–17 (2016).

69. Zhang, H. *et al.* TEAD Transcription Factors Mediate the Function of TAZ in Cell Growth and Epithelial-Mesenchymal Transition. *Journal of Biological Chemistry*

284, 13355–13362 (2009).

70. Rozengurt, E., Sinnett-Smith, J. & Eibl, G. Yes-associated protein (YAP) in pancreatic cancer: at the epicenter of a targetable signaling network associated with patient survival. *Signal Transduct Target Ther* **3**, 11 (2018).
71. Zhou, Y. *et al.* The TEAD Family and Its Oncogenic Role in Promoting Tumorigenesis. *Int. J. Mol. Sci.* **17**, (2016).
72. Yu, M.-H. & Zhang, W. TEAD1 enhances proliferation via activating SP1 in colorectal cancer. *Biomedicine & Pharmacotherapy* **83**, 496–501 (2016).
73. Liu, Y. *et al.* Increased TEAD4 expression and nuclear localization in colorectal cancer promote epithelial–mesenchymal transition and metastasis in a YAP-independent manner. *Oncogene* **35**, 2789 (2015).
74. Wang, C. *et al.* The interplay between TEAD4 and KLF5 promotes breast cancer partially through inhibiting the transcription of $p27^{Kip1}$. *Oncotarget* **6**, (2015).
75. Møller, M. B. p27 in Cell Cycle Control and Cancer. *Leukemia & Lymphoma* **39**, 19–27 (2000).
76. Zhao, B. *et al.* TEAD mediates YAP-dependent gene induction and growth control. *Genes Dev.* **22**, 1962–1971 (2008).
77. Xu, M. Z. *et al.* AXL receptor kinase is a mediator of YAP-dependent oncogenic functions in hepatocellular carcinoma. *Oncogene* **30**, 1229–1240 (2011).
78. Lai, D., Ho, K. C., Hao, Y. & Yang, X. Taxol Resistance in Breast Cancer Cells Is Mediated by the Hippo Pathway Component TAZ and Its Downstream Transcriptional Targets Cyr61 and CTGF. *Cancer Research* **71**, 2728–2738 (2011).
79. Knight, J. F. *et al.* TEAD1 and c-Cbl are novel prostate basal cell markers that

correlate with poor clinical outcome in prostate cancer. *Br. J. Cancer* **99**, 1849–1858 (2008).

80. Glück, S. *et al.* TP53 genomics predict higher clinical and pathologic tumor response in operable early-stage breast cancer treated with docetaxel-capecitabine ± trastuzumab. *Breast Cancer Research and Treatment* **132**, 781–791 (2012).
81. Blaveri, E. *et al.* Bladder Cancer Outcome and Subtype Classification by Gene Expression. *Journal of Urology* **175**, 1273–1273 (2006).
82. Blaveri, E. *et al.* Bladder Cancer Stage and Outcome by Array-Based Comparative Genomic Hybridization. *Journal of Urology* **175**, 1568–1569 (2006).
83. Zhou, Y. *et al.* TEAD1/4 exerts oncogenic role and is negatively regulated by miR-4269 in gastric tumorigenesis. *Oncogene* **36**, 6518–6530 (2017).
84. Pobbati, A. V. *et al.* Targeting the Central Pocket in Human Transcription Factor TEAD as a Potential Cancer Therapeutic Strategy. *Structure* **23**, 2076–2086 (2015).
85. Gibault, F., Sturbaut, M., Bailly, F., Melnyk, P. & Cotelle, P. Targeting Transcriptional Enhanced Associate Domains (TEADs). *Journal of Medicinal Chemistry* **61**, 5057–5072 (2018).
86. Benhaddou, A. *et al.* Transcription factor TEAD4 regulates expression of myogenin and the unfolded protein response genes during C2C12 cell differentiation. *Cell Death Differ.* **19**, 220–231 (2012).
87. Stewart, A. F. R. *et al.* Cloning of Human RTEF-1, a Transcriptional Enhancer Factor-1-Related Gene Preferentially Expressed in Skeletal Muscle: Evidence for an Ancient Multigene Family. *Genomics* **37**, 68–76 (1996).

88. Yockey, C. E., Smith, G., Izumo, S. & Shimizu, N. cDNA Cloning and Characterization of Murine Transcriptional Enhancer Factor-1-related Protein 1, a Transcription Factor That Binds to the M-CAT Motif. *Journal of Biological Chemistry* **271**, 3727–3736 (1996).
89. Nishioka, N. *et al.* Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech. Dev.* **125**, 270–283 (2008).
90. Yu, F.-X., Zhao, B. & Guan, K.-L. Hippo Pathway in Organ Size Control, Tissue Homeostasis, and Cancer. *Cell* **163**, 811–828 (2015).
91. Qi, Y. *et al.* A splicing isoform of TEAD4 attenuates the Hippo–YAP signalling to inhibit tumour proliferation. *Nature Communications* **7**, (2016).
92. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
93. Quagliata, L. *et al.* SH2D4A is frequently downregulated in hepatocellular carcinoma and cirrhotic nodules. *Eur. J. Cancer* **50**, 731–738 (2014).
94. Website. Available at: https://doi.org/10.1111/j.1365-2559.2011.04121. (Accessed: 11th February 2019)
95. Andreozzi, M. *et al.* HMGA1 Expression in Human Hepatocellular Carcinoma Correlates with Poor Prognosis and Promotes Tumor Growth and Migration in in vitro Models. *Neoplasia* **18**, 724–731 (2016).
96. Livak, K. J. & Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* **25**, 402–408 (2001).
97. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

98. Sergushichev, A. A. Algorithm for cumulative calculation of gene set enrichment statistic. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics* 956–959 (2016). doi:10.17586/2226-1494-2016-16-5-956-959

99. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems* **1**, 417–425 (2015).

100. Blecher-Gonen, R. *et al.* High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat. Protoc.* **8**, 539–554 (2013).

101. Dimitrova, Y. *et al.* TFAP2A is a component of the ZEB1/2 network that regulates TGFB1-induced epithelial to mesenchymal transition. *Biol. Direct* **12**, 8 (2017).

102. FireBrowse. Available at: http://firebrowse.org/. (Accessed: 11th April 2019)

103. Kancherla, V. *et al.* Genomic Analysis Revealed New Oncogenic Signatures in TP53-Mutant Hepatocellular Carcinoma. *Frontiers in Genetics* **9**, (2018).

104. Expression of TEAD4 in liver cancer - The Human Protein Atlas. Available at: https://www.proteinatlas.org/ENSG00000197905-TEAD4/pathology/tissue/liver+cancer. (Accessed: 2nd May 2019)

105. Website. Available at: https://cran.r-project.org/web/packages/maxstat/index.html]. (Accessed: 3rd May 2019)

106. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).

107. Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.*

46, D260–D266 (2018).

108. Jolly, C. Role of the Heat Shock Response and Molecular Chaperones in Oncogenesis and Cell Death. *Journal of the National Cancer Institute* **92**, 1564–1572 (2000).

109. Fuller, K. J. *et al.* Cancer and the heat shock response. *European Journal of Cancer* **30**, 1884–1891 (1994).

110. Seo, J. S. *et al.* T cell lymphoma in transgenic mice expressing the human Hsp70 gene. *Biochem. Biophys. Res. Commun.* **218**, 582–587 (1996).

111. Rohde, M. *et al.* Members of the heat-shock protein 70 family promote cancer cell growth by distinct mechanisms. *Genes Dev.* **19**, 570–582 (2005).

112. Lee, H. W. *et al.* Heat shock protein 70 (HSP70) expression is associated with poor prognosis in intestinal type gastric cancer. *Virchows Arch.* **463**, 489–495 (2013).

113. Syrigos, K. N. *et al.* Clinical significance of heat shock protein-70 expression in bladder cancer. *Urology* **61**, 677–680 (2003).

114. Barnes, J. A., Dix, D. J., Collins, B. W., Luft, C. & Allen, J. W. Expression of inducible Hsp70 enhances the proliferation of MCF-7 breast cancer cells and protects against the cytotoxic effects of hyperthermia. *Cell Stress Chaperones* **6**, 316–325 (2001).

115. Rérole, A.-L., Jego, G. & Garrido, C. Hsp70: anti-apoptotic and tumorigenic protein. *Methods Mol. Biol.* **787**, 205–230 (2011).

116. Sherman, M. Y. & Gabai, V. L. Hsp70 in cancer: back to the future. *Oncogene* **34**, 4153–4161 (2015).

117. Zanconato, F. *et al.* Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. *Nat. Cell Biol.* **17**, 1218–1227

(2015).

118. Xiong, J., Jiang, X.-M., Mao, S.-S., Yu, X.-N. & Huang, X.-X. Heat shock protein 70 downregulation inhibits proliferation, migration and tumorigenicity in hepatocellular carcinoma cells. *Oncol. Lett.* **14**, 2703–2708 (2017).

119. Liu, X. *et al.* Tead and AP1 Coordinate Transcription and Motility. *Cell Rep.* **14**, 1169–1180 (2016).

120. Hoshida, Y. *et al.* Molecular classification and novel targets in hepatocellular carcinoma: recent advancements. *Semin. Liver Dis.* **30**, 35–51 (2010).

121. Tang, J.-Y. *et al.* TEAD4 promotes colorectal tumorigenesis via transcriptionally targeting YAP1. *Cell Cycle* **17**, 102–109 (2018).

122. Zhang, Q. *et al.* TEAD4 exerts pro-metastatic effects and is negatively regulated by miR6839-3p in lung adenocarcinoma progression. *J. Cell. Mol. Med.* **22**, 3560–3571 (2018).

123. Martin, J. & -Ulrich Hartl, F. Molecular chaperones in cellular protein folding. *BioEssays* **16**, 689–692 (1994).

124. Ciocca, D. R. *et al.* Heat shock protein hsp70 in patients with axillary lymph node-negative breast cancer: prognostic implications. *J. Natl. Cancer Inst.* **85**, 570–574 (1993).

125. Hwang, T. S. *et al.* Differential, stage-dependent expression of Hsp70, Hsp110 and Bcl-2 in colorectal cancer. *J. Gastroenterol. Hepatol.* **18**, 690–700 (2003).

126. Joo, M., Chi, J. G. & Lee, H. Expressions of HSP70 and HSP27 in hepatocellular carcinoma. *J. Korean Med. Sci.* **20**, 829–834 (2005).

127. Tang, D. *et al.* Expression of heat shock proteins and heat shock protein messenger ribonucleic acid in human prostate carcinoma in vitro and in tumors in vivo. *Cell Stress Chaperones* **10**, 46–58 (2005).

128. Chuma, M., Sakamoto, M., Hige, S., Asaka, M. & Hirohashi, S. 503 Expression profiling in multistage hepatocarcinogenesis: identification of HSP70 as a molecular marker of early hepatocellular carcinoma. *Hepatology* **38**, 402–402 (2003).

129. Tommaso, L. D. *et al.* Diagnostic value of HSP70, glypican 3, and glutamine synthetase in hepatocellular nodules in cirrhosis. *Hepatology* **45**, 725–734 (2007).

9. APPENDIX

9.1 Diagnostic Targeted Sequencing Panel for Hepatocellular Carcinoma Genomic Screening.

2018 JMD CME Program

The Journal of Molecular Diagnostics, Vol. 20, No. 6, November 2018

ELSEVIER

the Journal of Molecular Diagnostics

jmd.amjpathol.org

Diagnostic Targeted Sequencing Panel for Hepatocellular Carcinoma Genomic Screening

Viola Paradiso,* Andrea Garofoli,* Nadia Tosti,* Manuela Lanzafame,* Valeria Perrina,* Luca Quagliata,* Matthias S. Matter,* Stefan Wieland,† Markus H. Heim,†‡ Salvatore Piscuoglio,* Charlotte K.Y. Ng,*† and Luigi M. Terracciano*

From the Institute of Pathology, University Hospital Basel, Basel; the Department of Biomedicine,† University of Basel, Basel; and the Department of Gastroenterology and Hepatology,‡ University Hospital Basel, Basel, Switzerland*

CME Accreditation Statement: This activity ("JMD 2018 CME Program in Molecular Diagnostics") has been planned and implemented in accordance with the accreditation requirements and policies of the Accreditation Council for Continuing Medical Education (ACCME) through the joint providership of the American Society for Clinical Pathology (ASCP) and the American Society for Investigative Pathology (ASIP). ASCP is accredited by the ACCME to provide continuing medical education for physicians.

The ASCP designates this journal-based CME activity ("JMD 2018 CME Program in Molecular Diagnostics") for a maximum of 18.0 AMA PRA Category 1 Credit(s)™. Physicians should claim only credit commensurate with the extent of their participation in the activity.

CME Disclosures: The authors of this article and the planning committee members and staff have no relevant financial relationships with commercial interests to disclose.

Accepted for publication July 2, 2018.

Address correspondence to Luigi M. Terracciano, M.D., Institute of Pathology, University Hospital Basel, Schoenbeinstrasse 40, 4031 Basel, Switzerland. E-mail: luigi.terracciano@usb.ch.

Commercially available targeted panels miss genomic regions frequently altered in hepatocellular carcinoma (HCC). We sought to design and benchmark a sequencing assay for genomic screening of HCC. We designed an AmpliSeq custom panel targeting all exons of 33 protein-coding and two long non-coding RNA genes frequently mutated in HCC, *TERT* promoter, and nine genes with frequent copy number alterations. By using this panel, the profiling of DNA from fresh-frozen ($n = 10$, 1495×) and/or formalin-fixed, paraffin-embedded (FFPE) tumors with low-input DNA ($n = 36$, 530×) from 39 HCCs identified at least one somatic mutation in 90% of the cases. Median of 2.5 (range, 0 to 74) and 3 (range, 0 to 76) mutations were identified in fresh-frozen and FFPE tumors, respectively. Benchmarked against the mutations identified from Illumina whole-exome sequencing (WES) of the corresponding fresh-frozen tumors (105×), 98% (61 of 62) and 100% (104 of 104) of the mutations from WES were detected in the 10 fresh-frozen tumors and the 36 FFPE tumors, respectively, using the HCC panel. In addition, 18 and 70 somatic mutations in coding and noncoding genes, respectively, not found by WES were identified by using our HCC panel. Copy number alterations between WES and our HCC panel showed an overall concordance of 86%. In conclusion, we established a cost-effective assay for the detection of genomic alterations in HCC. *(J Mol Diagn 2018, 20: 836—848; https://doi.org/10.1016/j.jmoldx.2018.07.003)*

Sequencing technologies have allowed the discovery of genetic alterations essential in the diagnosis and treatment of human cancer or approval of new targeted therapies.[1] In addition, the presence of subclonal mutations has direct implications in the development of drug resistance.[2,3] In the era of precision medicine, the development of rapid, accurate, high-throughput, and cost-effective genomic assays to accommodate the increasingly genotype-based therapeutic approaches is required.[4,5] Currently, the costs of whole-genome and whole-exome sequencing (WES) are still prohibitive in the clinical setting, especially for small institutions. Furthermore, although DNA from fresh-frozen

Supported in part by the Swiss Cancer League (Oncosuisse) grants KLS-3639-02-2015 (L.M.T.) and KFS-3995-08-2016 (S.P.), Krebsliga beider Basel project KLbB-4183-03-2017 (C.K.Y.N.), Swiss National Science Foundation Ambizione grant PZ00P3_168165 (S.P.), the Swiss Centre for Applied Human Toxicology (SCAHT; V.Pa.), and the European Research Council ERC Synergy grant 609883 (C.K.Y.N. and M.H.H.).

V.Pa. and A.G. contributed equally to this work.

C.K.Y.N. and L.M.T. contributed equally to this work as senior authors.

Disclosures: None declared.

Funding bodies had no role in the design of the study, collection, analysis, and interpretation of the data or the writing of the manuscript.

https://doi.org/10.1016/j.jmoldx.2018.07.003

tissue is ideal for genomic screening, it is not part of routine diagnostic practice at most hospitals and institutions. Instead, DNA from formalin-fixed paraffin-embedded (FFPE) material is frequently the only option. Moreover, DNA from small tumors, after reserving materials for histopathologic analyses, may be extremely limited. For research institutes, being able to exploit and revisit archival materials associated with long-term follow-up but whose DNA may potentially be degraded is also highly desirable. Given these limitations, PCR-based sequencing panels may be more broadly applicable than capture-based solutions.

Existing commercial sequencing panels, such as the amplicon-based Ion Torrent Oncomine Comprehensive Assay version 3 (Thermo Fisher Scientific, Waltham, MA) and the capture-based Foundation Medicine FoundationOne assay, are broadly applicable to common cancer types. Compared with other common cancer types, however, hepatocellular carcinoma (HCC) has a distinct mutational profile. Although HCC driver genes *TP53* and *CTNNB1* are also frequently mutated in cancers such as those of the lungs, the breasts, and colon,[6] genes such as *APOB*, *ALB*, *HNF1A*, and *HNF4A* are significantly mutated only in HCC.[7–17] The distinct mutational landscape of HCC is likely a result of the unique biology of hepatocyte differentiation and liver functions. Of note, the frequently altered *APOB*, *ALB*, and *HNF4A* are not targeted by most commercial assays. In the noncoding regions, recent commercially available panels include *TERT* promoter mutation hotspot (c.-124C>T). However, long noncoding RNA (lncRNA) genes frequently mutated in HCC, such as *MALAT1* and *NEAT1*,[16] have yet to be included in commercial panels or in exome capture panels. Recent whole-genome studies have also uncovered mutation clusters in promoter regions of genes such as *MED16*, *WDR74*, and *TFPI2*[16,18] that are not covered in commercial panels.

In this study, we designed a high-throughput and cost-effective amplicon-based sequencing panel specifically to screen for somatic mutations and copy number alterations (CNAs) in HCC. Our panel includes genes and regions frequently altered in HCC, including those not currently covered by commercial panels. We tested the sequencing panel by using fresh-frozen and FFPE materials with low-input DNA to evaluate the feasibility of this panel in routine diagnostics.

Materials and Methods

Targeted Panel Design and Generation

A custom targeted sequencing panel that focused on the most frequently altered genes in HCC[7–18] was designed by using Ion Ampliseq Designer (Thermo Fisher Scientific). The panel (hereafter the HCC panel) covers all exons of 33 protein-coding genes; recurrently mutated lncRNA genes *MALAT1* and *NEAT1*; and the recurrently mutated promoter regions of *TERT*, *WDR74*, *MED16*, and *TFPI2* (Figure 1A and Supplemental Table S1).[7–18] Nine genes frequently altered by CNAs and mutation hotspots in seven cancer genes are also covered (Figure 1A and Supplemental Table S1).[7–18] The HCC panel was designed by using the FFPE option for smaller amplicon size. The nine genes for CNA profiling were designed to be covered by at least 10 nonoverlapping amplicons evenly distributed across the length of the genes. The designed panel was further inspected by

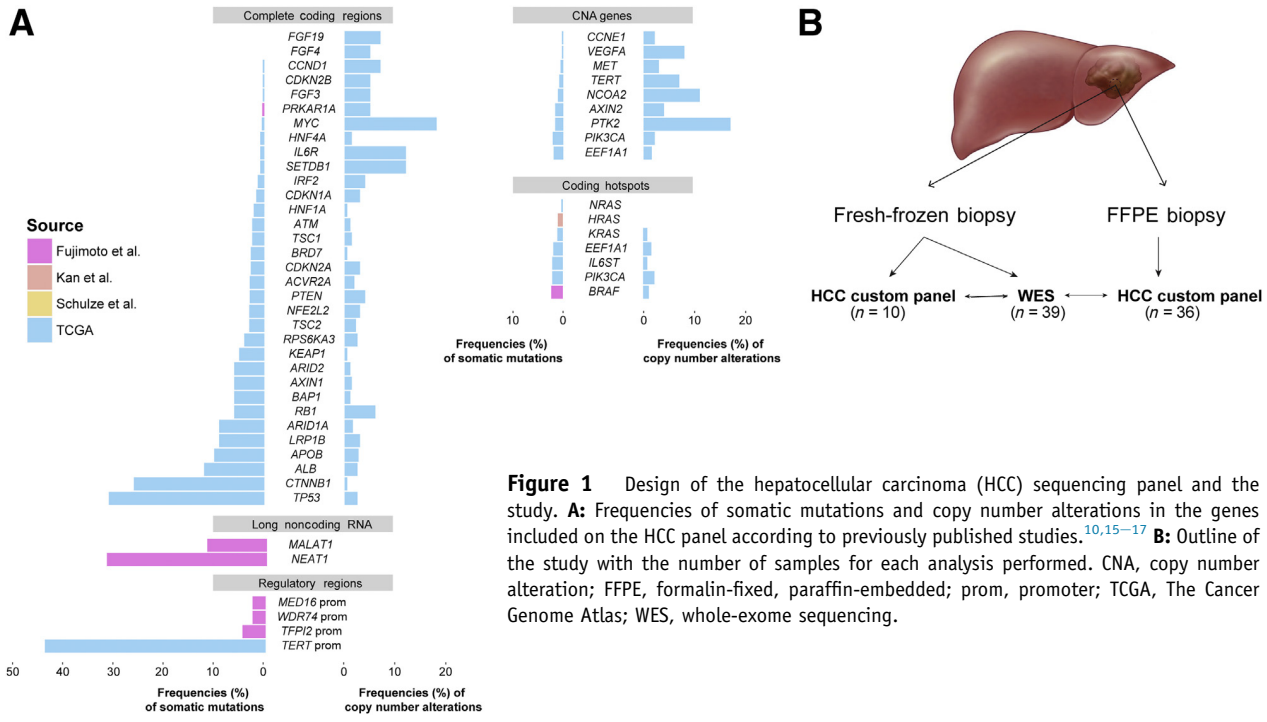


Figure 1 Design of the hepatocellular carcinoma (HCC) sequencing panel and the study. **A:** Frequencies of somatic mutations and copy number alterations in the genes included on the HCC panel according to previously published studies.[10,15–17] **B:** Outline of the study with the number of samples for each analysis performed. CNA, copy number alteration; FFPE, formalin-fixed, paraffin-embedded; prom, promoter; TCGA, The Cancer Genome Atlas; WES, whole-exome sequencing.

the white glove service (Thermo Fisher Scientific) for primer specificity in a multiplex PCR reaction. The HCC panel consists of 2120 amplicons split into two primer pools and covers genomic regions of approximately 203 kb.

Tissue Samples

Human tissues were obtained from patients undergoing diagnostic liver biopsy at the University Hospital Basel, Basel, Switzerland. Written informed consent was obtained from all included patients. Ultrasound-guided needle biopsies were obtained from tumor lesion(s) and adjacent nontumoral liver tissue (Figure 1B). The study was approved by the ethics committee of the northwestern part of Switzerland (protocol EKNZ 2014-099). For all patients except cases 2, 6, 7, and 9, a single tumor biopsy was included (Supplemental Table S2). For cases 6 and 7, two tumor biopsies were included, and for cases 2 and 9, three tumor biopsies were included. A portion of each biopsy was FFPE for clinical purposes, and the remaining portion of each biopsy was snap-frozen and stored at −80°C for research purposes. For this study, 45 fresh-frozen tumor biopsies and 39 fresh-frozen nontumor biopsies from 39 patients were included. FFPE tissue samples that remained after diagnostic routine (36 tumor biopsies and 31 nontumor biopsies from 36 patients) were included. Pathologic assessment of tumor content was performed by two expert hepatopathologists (M.S.M. and L.M.T.) with the use of diagnostic hematoxylin and eosin slides.

DNA Extraction

DNA from fresh-frozen biopsies was extracted by using the ZR-Duet DNA/RNA MiniPrep Plus kit (Zymo Research, Irvine, CA) according to the manufacturer's instructions. Before extraction, tissue samples were crushed in liquid nitrogen to facilitate lysis. For DNA extraction from FFPE samples, one 5-μm−thick slide was cut directly in the tube, and DNA was extracted with the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions as previously described.[19,20] DNA was quantified by using the Qubit Fluorometer (Thermo Fisher Scientific).

Library Preparation and Deep Sequencing Using the HCC Panel

Library preparation for the HCC panel was performed by using the Ion AmpliSeq library kit version 2.0 (Thermo Fisher Scientific) according to the manufacturer's guidelines. For cases 2, 6, 7, and 9, DNA extracted from multiple fresh-frozen tumor biopsies was pooled equimolar before library preparation (Supplemental Table S2). In total, 20 fresh-frozen samples (10 tumor samples and 10 nontumoral counterparts) and 67 FFPE samples (36 tumor biopsies and 31 nontumoral counterparts) were sequenced by using the HCC panel.

The HCC panel consists of two pools of amplification primers. Ten nanograms of DNA per sample was used for library preparation for each pool. Amplification was performed according to the manufacturer's guidelines. The amplicons from the two pools were combined and treated to digest the primers and to phosphorylate the amplicons. The amplicons were then ligated to Ion Adapters (Thermo Fisher Scientific) by using DNA ligase. Finally, cleaning and purification of the generated libraries were performed with Agencourt AMPure XP (Beckman Coulter, Brea, CA) according to the manufacturer's guidelines. Quantification and quality control were performed with the Ion Library TaqMan Quantitation Kit (Thermo Fisher Scientific). Samples were diluted to reach the concentration of 40 pmol and then were pooled for sequencing. Twenty-five μL of the pooled libraries was loaded on Ion 530 Chip (Thermo Fisher Scientific) and processed in Ion Chef Instrument (Thermo Fisher Scientific). Sequencing was performed on Ion S5 XL system (Thermo Fisher Scientific).

Sequence Data Analysis for the HCC Panel

Sequence reads were aligned to the human reference genome hg19 by using TMAP within the Torrent Suite Software version 5.4 (Thermo Fisher Scientific; *https://github.com/iontorrent/TS*) for the Ion S5XL system. Coverage analysis was performed by using Picard's CollectTargetedPcrMetrics tool version 2.4.1 (*http://broadinstitute.github.io/picard*) (Supplemental Table S3). Uniformity of sequencing was defined as the proportion of target bases covered at >20% of mean amplicon coverage for a given sample. Comparison of the coverage for the two primer pools was performed by using paired Wilcoxon test.

Somatic mutations were identified with Torrent Variant Caller version 5.0.3 (Thermo Fisher Scientific; *https://github.com/iontorrent/TS*). For fresh-frozen samples, the corresponding fresh-frozen nontumoral samples were used as the germline control. For FFPE samples, FFPE nontumoral samples were used as the matched germline sample when available. When FFPE nontumoral samples were not available, the corresponding fresh-frozen nontumoral samples were used as germline control. Mutations at hotspot residues were whitelisted.[21,22] Mutations supported by <8 reads, and/or those covered by <10 reads in the tumor or <10 reads in the matched nontumoral counterpart were filtered out. Only those for which the tumor variant allele fraction (VAF) was >10 times that of the matched nontumoral VAF were retained to ensure the somatic nature of the variants. Because of the repetitive nature and the high GC content of the *TERT* promoter region, *TERT* mutation hotspots (chr5:1295228 and chr5:1295250) were additionally screened. *TERT* promoter mutations were considered present if supported by at least five reads or VAF of at least 5%. All mutations were manually inspected by using the Integrative Genomics Viewer version 2.3.69 (*https://software.broadinstitute.org/software/igv*).[23]

CNAs were defined as follows. For each sample, end-to-end sequence reads were extracted separately for the two

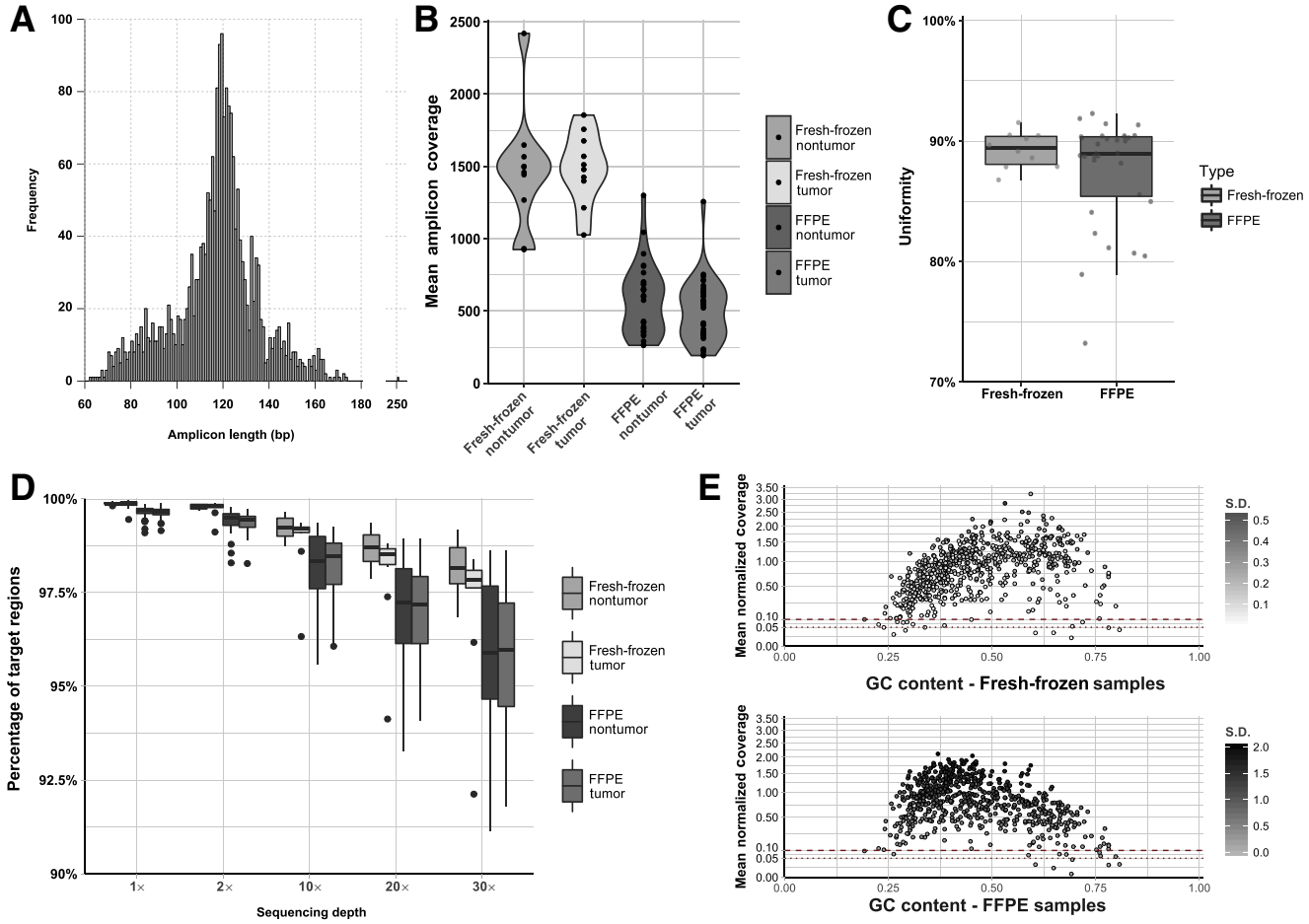


Figure 2 Coverage analyses and statistics of the hepatocellular carcinoma (HCC) panel. **A:** Distribution of the amplicon sizes on the HCC panel. **B:** Violin plots of the mean amplicon coverage across fresh-frozen nontumor; fresh-frozen tumor; formalin-fixed, paraffin-embedded (FFPE) nontumor; and FFPE tumor samples. **C:** Coverage uniformity, defined as the percentage of target bases covered at >20% of the mean coverage, in fresh-frozen and FFPE nontumor samples. **D:** Percentages of target regions covered at various depths (1×, 2×, 10×, 20×, and 30×) across fresh-frozen nontumor, fresh-frozen tumor, FFPE nontumor, and FFPE tumor samples. **E:** Scatter plot of GC content and mean normalized coverage for all amplicons in fresh-frozen and FFPE samples. Color of the dots indicates the SD of mean normalized coverage within each group. **Dashed red lines** indicate the mean normalized coverage at 0.1 and 0.05.

amplicon pools. A copy number reference for each pool was generated by using all nontumoral samples to estimate overall read depth, $\log_2$ ratio, and variability by using the reference function from CNVkit version 0.9.0 (*https://github.com/etal/cnvkit*).[24] Amplicons with <100 read depth, absolute $\log_2$ ratio >1.5, or spread >1 were removed from copy number analysis. Protein-coding genes for which the complete coding region was included in the panel or for which amplicons were specifically designed for copy number analysis were included. Samples with excessive residual copy number $\log_2$ ratio (segment interquartile range >0.8) were excluded, as previously described.[25]

For each tumor/nontumor pair, $\log_2$ ratio was computed for each amplicon, separately for the two amplicon pools by using Varscan2 version 2.4.3 (*https://github.com/dkoboldt/varscan*).[26] Log_2 ratios for the two pools were separately centered then merged for segmentation by using circular binary segmentation.[27] CNAs were determined, adopting a previously described approach.[20] In brief, SD of the $\log_2$ ratios of the 40% of the central positions ordered by their $\log_2$ ratios was computed. Copy number gains and amplifications/high gains were defined as +2 SDs and +6 SDs, respectively. Copy number losses and deep deletions were defined as −2.5 SDs and −7 SDs, respectively. All gene amplifications and deep deletions were visually inspected by using $\log_2$ ratio plots.

To evaluate the impact of tumor purity on CNA analysis, an *in silico* simulation was performed on 12 cases (six frozen and six FFPE, selected on the basis of the presence of gene amplification/high gain or deep deletion), by replacing tumor reads with reads sampled from the normal samples to simulate tumor content 5%, 10%, 20% up to the actual tumor content for the samples. CNA analysis was performed as described above.

WES

WES was performed for DNA extracted from the 45 tumor biopsies and 39 nontumoral counterparts from the 39

patients (Supplemental Table S2). Whole-exome capture was performed by using the SureSelectXT Clinical Research Exome (Agilent, Santa Clara, CA) platform according to the manufacturer's guidelines. Sequencing (2 × 101 bp) was performed at the Genomics Facility of ETH Zurich Department of Biosystems Science and Engineering (Basel, Switzerland) by using Illumina HiSeq 2500 (Illumina, San Diego, CA) according to the manufacturer's guidelines. Sequence reads were aligned to the reference human genome GRCh37 by using Burrows-Wheeler Aligner-MEM version 0.7.12 (*http://bio-bwa.sourceforge.net*).[28] Local realignment, duplicate removal, and base quality adjustment were performed by using the Genome Analysis Toolkit version 3.6 (*https://software.broadinstitute.org/gatk*)[29] and Picard version 2.4.1 (*http://broadinstitute.github.io/picard*).

For WES samples, sequence reads overlapping with the target regions of the HCC panel were extracted for further comparative analyses. Sequencing statistics were evaluated for the overlap of the target regions of the WES and the HCC panel. For cases 2, 6, 7, and 9, for which DNA from multiple fresh-frozen tumor biopsies was pooled before sequencing by using the HCC panel, WES reads from the multiple biopsies were merged to facilitate downstream comparisons. For all four cases, the number of reads obtained from WES of individual biopsies was comparable (Supplemental Table S3).

Somatic single nucleotide variants and small insertions and deletions (indels) were detected by using MuTect version 1.1.4 (*https://software.broadinstitute.org/cancer/cga/mutect*)[30] and Strelka version 1.0.15 (*https://github.com/Illumina/strelka*),[31] respectively. Single nucleotide variants and small indels outside of the target regions, those with VAF of <1%, and/or those supported by <3 reads were filtered out. Only variants for which the tumor VAF was >5 times that of the matched nontumoral VAF were retained. Further, variants identified in at least two of a panel of 123 nontumoral liver tissue samples, using the artifact detection mode of MuTect2 implemented in Genome Analysis Toolkit version 3.6 were excluded,[29] where the panel of 123 nontumoral liver tissue samples included the 39 nontumoral samples in the present study and were captured and sequenced with the same protocols. All indels were manually inspected by using the Integrative Genomics Viewer.[23] Copy number analysis was performed with FACETS version 0.5.13 (*https://github.com/mskcc/facets*),[32] and genes targeted by amplifications or deep deletions were defined by using the same thresholds as above.

Pairwise Comparisons between Mutations Identified by WES, Fresh-Frozen and FFPE Tissues

Pairwise comparisons of the somatic mutations identified by WES and by the HCC panel were performed, according to the originating biopsies (Supplemental Table S2). Discordant variants were reevaluated and interrogated for their presence by supplying Torrent Variant Caller version 5.0.3 with their positions as the hotspot list (for Ion Torrent sequencing) or by Genome Analysis Toolkit version 3.6 Unified Genotyper by using the GENOTYPE_GIVEN_ALLELES mode (for WES).

Sanger Sequencing

To validate the discordant variants, Sanger sequencing was performed on both DNA from the fresh-frozen and the corresponding FFPE tumor biopsies. PCR amplification of 5 ng of genomic DNA was performed with the AmpliTaq 360 Master Mix Kit (Thermo Fisher Scientific) on a Veriti Thermal Cycler (Thermo Fisher Scientific) as previously described (Supplemental Table S4).[20] PCR fragments were purified with ExoSAP-IT (Thermo Fisher Scientific). Sequencing reactions were performed on a 3500 Series Genetic Analyzer instrument by using the ABI BigDye Terminator chemistry version 3.1 (Thermo Fisher Scientific) according to the manufacturer's instructions. All analyses were performed in duplicate. Sequences of the forward and reverse strands were analyzed with MacVector software version 15.1.3 (MacVector, Inc., Apex, NC).[20]

Analysis of TCGA Data

To determine the frequencies of high-level copy number gains/focal amplifications and deep deletions/focal homozygous deletions in HCC, the GISTIC 2.0 copy number calls for The Cancer Genome Atlas (TCGA) HCC cohort from the cBioPortal were obtained.[33] High-level gains and deep deletions were defined as those with GISTIC copy number state 2 and −2, respectively. Focal amplifications and focal homozygous deletions were defined as high-level gains and deep deletions that affected <25% of a given chromosome arm. For the 37 genes included in the copy number analysis, the frequencies of high-level gains/deep deletions and of focal amplifications/focal homozygous deletions were computed.

Statistical Analysis

Correlation analyses were performed with Pearson's r and r^2. Statistical analyses were performed in R version 3.4.2 (The R Foundation, Vienna, Austria).

Results

HCC-Specific Custom Targeted Sequencing Panel Design and Quality Assessment

An HCC sequencing panel was designed to specifically target genes and genomic regions frequently altered in HCC[7–18] (Figure 1A and Supplemental Table S1). The HCC panel consisted of complete coding regions of 33 genes involved in several pathways implicated in HCC pathogenesis, including the WNT pathway (*CTNNB1*, *AXIN1*), chromatin remodeling (*ARID1A*, *ARID2*, and *BAP1*), cell cycle regulation (*CDKN1A*, *CDKN2A*, *CDKN2B*, *CCND1*, *RPS6KA3*, *RB1*, and *TP53*),

inflammatory response (*IL6R*, *IL6ST*), and hepatocyte differentiation (*ALB*, *APOB*, *HNF1A*, and *HNF4A*). In addition, the HCC panel also targeted recurrently mutated lncRNA genes *MALAT1* and *NEAT1* and recurrently mutated promoter regions of *TERT*, *WDR74*, *MED16*, and *TFPI2*. Genes frequently altered by CNAs (eg, *CCNE1*, *VEGFA*, *TERT*) and mutation hotspots in *BRAF*, *EEF1A1*, *HRAS*, *IL6ST*, *KRAS*, *NRAS*, and *PIK3CA* were also targeted. To enable the efficient profiling of DNA samples derived from potentially degraded FFPE materials, the panel was designed by using the FFPE option for smaller amplicon size, with a mean amplicon size of 118 bp (range, 63 to 252 bp) (Figure 2A). The HCC panel was tested on the DNA extracted from 20 fresh-frozen samples (10 from tumor biopsies and 10 from nontumoral counterparts) and 67 FFPE samples (36 from tumor biopsies and 31 from nontumoral counterparts) obtained from 39 patients (Figure 1B and Supplemental Table S2).

A coverage analysis of the HCC panel was performed with the 10 fresh-frozen and 31 FFPE nontumoral DNA samples. In the fresh-frozen and FFPE nontumoral DNA samples, a mean coverage of 1478× (range, 925× to 2420×) and 580× (range, 263× to 1300×), respectively, were achieved (Figure 2B and Supplemental Table S3). No difference was found between the depth of coverage of the two pools of amplicons ($P = 0.9879$, paired Wilcoxon test) (Supplemental Figure S1A). At least 96.8% and 91.1% of the amplicons were covered at >30× and at least 98.7% and 95.6% of the amplicons were covered at >10× in the fresh-frozen and FFPE nontumor samples, respectively (Figure 2C and Supplemental Figure S1B). Median uniformity (defined as the proportion of target bases covered at >20% of the mean amplicon coverage of a given sample) was 89.9% (range, 86.8% to 91.5%) in the fresh-frozen samples and 89.0% (range, 73.3% to 92.3%) in the FFPE samples (Figure 2D). As expected, depth of sequencing of the amplicons was associated with GC content, with reduced depth at extreme GC content (Figure 2E).

HCC Panel Captures Somatic Mutations Concordant with WES and Identifies Additional Mutations

Next, the somatic mutations identified in the 10 fresh-frozen tumor/nontumoral pairs sequenced with the HCC panel were evaluated. A median sequencing depth of 1495× (range, 1026× to 1855×) in the tumor samples was achieved (Figure 2B and Supplemental Table S3). A median of 2.5 somatic mutations (range, 0 to 74 somatic mutations) were identified, including a median of 2 mutations (range, 0 to 52 mutation) in protein-coding genes (Figure 3A and Supplemental Table S4). No somatic mutations were identified for 2 of 10 cases (cases 3 and 12), although both cases had ≥50% tumor cell content (Supplemental Table S2). One case (case 9) exhibited a hypermutator phenotype with 74 somatic mutations identified.

To evaluate the somatic mutations defined with the HCC panel, the somatic mutations derived from WES, generated on the orthogonal Illumina technology, of the same DNA aliquots from the fresh-frozen tumors and matched nontumor samples were used as a benchmark (Figure 1B). By considering only the coding regions covered by the HCC panel, the median depths of WES were 114× (range, 92× to 345×) and 51× (range, 45× to 84×) in the fresh-frozen tumors and matched nontumor samples, respectively (Supplemental Table S3). WES analysis confirmed that no mutations were present within the targeted protein-coding regions in cases 3 and 12 and that case 9 was hypermutated (Figure 3B). Of the 62 mutations in the coding region identified from WES analysis, 61 (98%) were also called by the HCC panel analysis (Figure 3B). One *NRAS* Q61K hotspot mutation (case 6) was missed by using the HCC panel analysis. Manual review of this position revealed that the mutation had VAF of 2.5% by WES and 2.0% by the HCC panel (Supplemental Figure S2 and Supplemental Table S4). Note, however, that 2% is close to the detection limit of the current sequencing technologies.

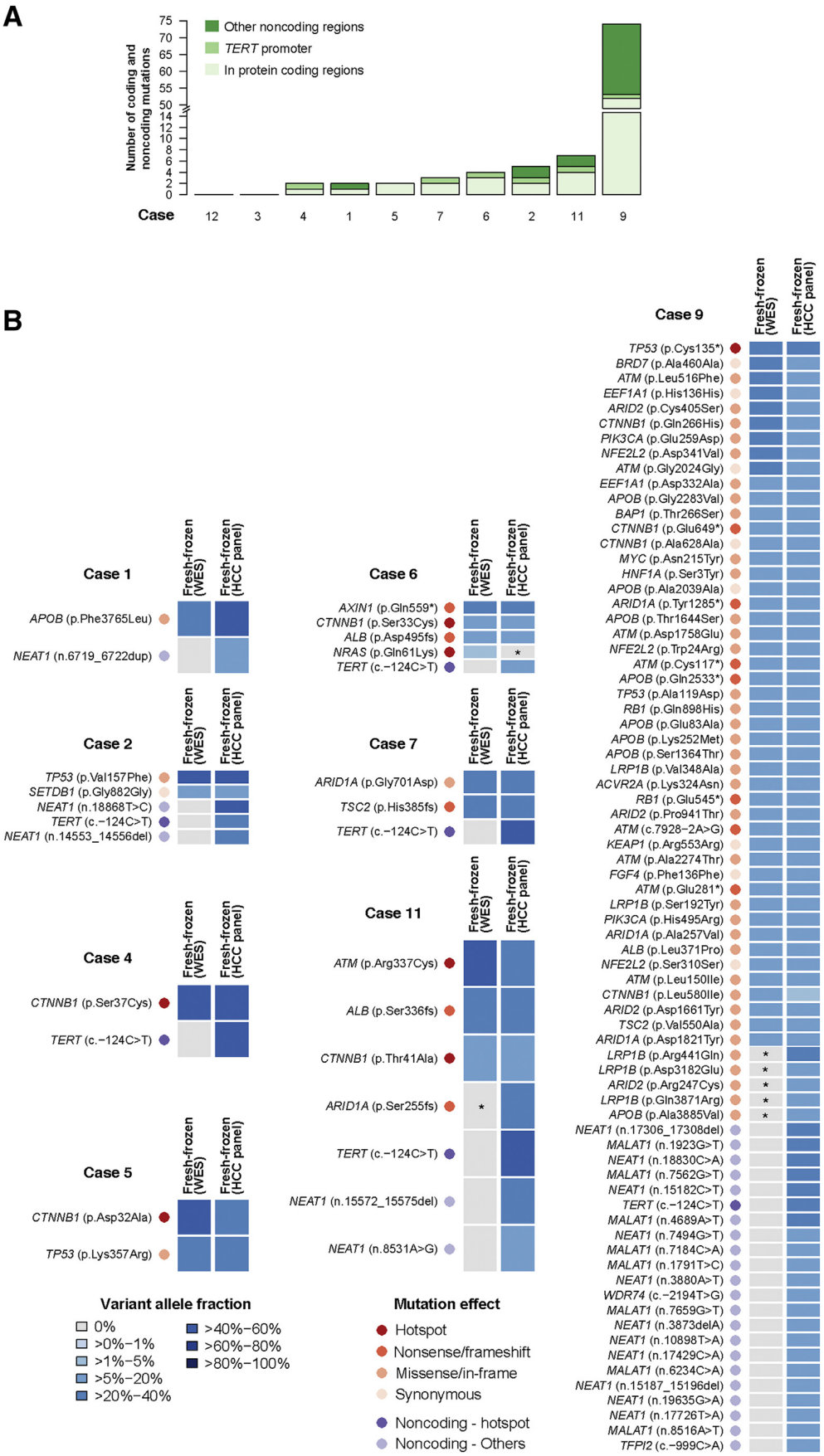
Compared with the WES analysis, the HCC panel analysis revealed an additional six mutations in the coding regions, including five in case 9 and one in case 11 (Figure 3B). Manual review of the WES data showed that all six mutations were in fact supported by at least one read in WES, but those positions were covered at reduced depth, with 4 of 6 covered by ≤40 reads (including three in *LRP1B*) and 5 of 6 ≤80 reads (Supplemental Figure S2C and Supplemental Table S4). This suggested that the increased sensitivity in the HCC panel analysis was likely due to the increased depth achieved.

Additional to the mutations in the protein-coding regions, the HCC panel also targeted the lncRNA genes *MALAT1* and *NEAT1* and the promoter regions of *TERT*, *WDR74*, *MED16*, and *TFPI2* (Figure 1A). Within these noncoding regions, an additional 32 mutations were identified across the 10 cases, representing a 48% gain of information compared with sequencing the protein-coding genes alone (Figure 3B). *TERT* promoter mutations were found in 60% (6 of 10) of cases and 16 somatic mutations in the lncRNA gene *NEAT1* were identified in 40% (4 of 10) of cases (Figure 3B and Supplemental Table S4).

Taken together, for the protein-coding genes frequently mutated in HCC, the HCC panel analysis produced highly reliable results compared with WES. Given the increased sequencing depth achieved by using the HCC panel, somatic mutations that were missed by WES were identified. Of importance, the HCC panel analysis enabled us to identify somatic mutations in promoter regions and frequently mutated lncRNA genes.

HCC Panel Analysis Identifies Somatic Mutations in FFPE Diagnostic Biopsies with Low-Input DNA

Nucleic acids from diagnostic specimens are frequently derived from small FFPE samples. Therefore, it would be important to determine whether the HCC panel could also be used for somatic mutational screening on low-input DNA


A
Number of coding and noncoding mutations
Other noncoding regions
TERT promoter
In protein coding regions
Case
12
3
4
1
5
7
6
2
11
9
B
Case 1
APOB (p.Phe3765Leu)
NEAT1 (n.6719_6722dup)
Case 2
TP53 (p.Val157Phe)
SETDB1 (p.Gly882Gly)
NEAT1 (n.18868T>C)
TERT (c.−124C>T)
NEAT1 (n.14553_14556del)
Case 4
CTNNB1 (p.Ser37Cys)
TERT (c.−124C>T)
Case 5
CTNNB1 (p.Asp32Ala)
TP53 (p.Lys357Arg)
Case 6
AXIN1 (p.Gln559*)
CTNNB1 (p.Ser33Cys)
ALB (p.Asp495fs)
NRAS (p.Gln61Lys)
TERT (c.−124C>T)
Case 7
ARID1A (p.Gly701Asp)
TSC2 (p.His385fs)
TERT (c.−124C>T)
Case 11
ATM (p.Arg337Cys)
ALB (p.Ser336fs)
CTNNB1 (p.Thr41Ala)
ARID1A (p.Ser255fs)
TERT (c.−124C>T)
NEAT1 (n.15572_15575del)
NEAT1 (n.8531A>G)
Fresh-frozen (WES)
Fresh-frozen (HCC panel)
Variant allele fraction
0%
>0%−1%
>1%−5%
>5%−20%
>20%−40%
>40%−60%
>60%−80%
>80%−100%
Mutation effect
Hotspot
Nonsense/frameshift
Missense/in-frame
Synonymous
Noncoding - hotspot
Noncoding - Others
Case 9
TP53 (p.Cys135*)
BRD7 (p.Ala460Ala)
ATM (p.Leu516Phe)
EEF1A1 (p.His136His)
ARID2 (p.Cys405Ser)
CTNNB1 (p.Gln266His)
PIK3CA (p.Glu259Asp)
NFE2L2 (p.Asp341Val)
ATM (p.Gly2024Gly)
EEF1A1 (p.Asp332Ala)
APOB (p.Gly2283Val)
BAP1 (p.Thr266Ser)
CTNNB1 (p.Glu649*)
CTNNB1 (p.Ala628Ala)
MYC (p.Asn215Tyr)
HNF1A (p.Ser3Tyr)
APOB (p.Ala2039Ala)
ARID1A (p.Tyr1285*)
APOB (p.Thr1644Ser)
ATM (p.Asp1758Glu)
NFE2L2 (p.Trp24Arg)
ATM (p.Cys117*)
APOB (p.Gln2533*)
TP53 (p.Ala119Asp)
RB1 (p.Gln898His)
APOB (p.Glu83Ala)
APOB (p.Lys252Met)
APOB (p.Ser1364Thr)
LRP1B (p.Val348Ala)
ACVR2A (p.Lys324Asn)
RB1 (p.Glu545*)
ARID2 (p.Pro941Thr)
ATM (c.7928−2A>G)
KEAP1 (p.Arg553Arg)
ATM (p.Ala2274Thr)
FGF4 (p.Phe136Phe)
ATM (p.Glu281*)
LRP1B (p.Ser192Tyr)
PIK3CA (p.His495Arg)
ARID1A (p.Ala257Val)
ALB (p.Leu371Pro)
NFE2L2 (p.Ser310Ser)
ATM (p.Leu150Ile)
CTNNB1 (p.Leu580Ile)
ARID2 (p.Asp1661Tyr)
TSC2 (p.Val550Ala)
ARID1A (p.Asp1821Tyr)
LRP1B (p.Arg441Gln)
LRP1B (p.Asp3182Glu)
ARID2 (p.Arg247Cys)
LRP1B (p.Gln3871Arg)
APOB (p.Ala3885Val)
NEAT1 (n.17306_17308del)
MALAT1 (n.1923G>T)
NEAT1 (n.18830C>A)
MALAT1 (n.7562G>T)
NEAT1 (n.15182C>T)
TERT (c.−124C>T)
MALAT1 (n.4689A>T)
NEAT1 (n.7494G>T)
MALAT1 (n.7184C>A)
MALAT1 (n.1791T>C)
NEAT1 (n.3880A>T)
WDR74 (c.−2194T>G)
MALAT1 (n.7659G>T)
NEAT1 (n.3873delA)
NEAT1 (n.10898T>A)
NEAT1 (n.17429C>A)
MALAT1 (n.6234C>A)
NEAT1 (n.15187_15196del)
NEAT1 (n.19635G>A)
NEAT1 (n.17726T>A)
MALAT1 (n.8516A>T)
TFPI2 (c.−999C>A)

(20 ng) extracted from FFPE samples. The DNA extracted from 36 diagnostic FFPE tumor biopsies was subjected to HCC panel sequencing to a median depth of 530× (range, 192× to 1257×) (Figures 1A and 2, B and C, and Supplemental Table S3). The median tumor content for these 36 cases was 90% (range, 5% to 100%) (Supplemental Table S2), thus representative of the distribution of tumor content in diagnostic samples in clinical practice. A median of three mutations (range, 0 to 76 mutations) per sample, including a median of two mutations (range, 0 to 53 mutations) in the coding regions was identified (Figure 4, Supplemental Figure S3, and Supplemental Table S4). No somatic mutations were identified for 8% (3 of 36) of cases (cases 7, 12, and 37), indicating that at least one somatic mutation could be detected in 92% of HCC diagnostic samples. Of note, although somatic mutations in the one biopsy with 5% tumor content could not be detected, somatic alterations in samples with 30% to 40% tumor content were detected.

The mutations identified in protein-coding genes from these 36 FFPE diagnostic biopsies were compared with those identified by WES of the DNA from the corresponding fresh-frozen biopsies. All 104 mutations identified from WES analysis were also called based on the HCC panel analysis (Figure 4 and Supplemental Figure S3), with 21 of 36 cases (58%) harboring *CTNNB1* mutations, a higher proportion than the TCGA and other HCC cohorts that was likely due to the higher percentage of alcohol-associated HCC (Supplemental Tables S1 and S2).[15] In addition, analysis of the HCC panel identified 18 mutations in the coding regions that were not found in the WES analysis in 11 cases. Of these 18 mutations, 13 were evident in WES but were not identified as mutations in the WES analysis, predominantly because of low sequencing depth (Supplemental Figures S2D and S3). The remaining five mutations were verified to be present in the corresponding FFPE samples but absent in the fresh-frozen samples by Sanger sequencing (Supplemental Figure S4 and Supplemental Table S4), indicating that they were genuine discordances between the fresh-frozen and FFPE DNA and not false positive calls from the HCC panel assay. Of note, two of five mutations validated to be absent from the fresh-frozen DNA affected mutation hotspots in *CTNNB1* (D32N and S45A) (Figure 4 and Supplemental Figure S4). The increased number of detected mutations by the HCC panel analysis was likely due to a combination of intratumor heterogeneity and the higher sequencing depth achieved.

Considering the 36 FFPE diagnostic biopsies, the HCC panel identified 70 somatic mutations in lncRNA genes and promoter regions, including 22 *TERT* promoter mutations (Figure 4 and Supplemental Table S4). Somatic mutations in lncRNA genes and promoter regions accounted for 37% of the total number of somatic mutations identified in the FFPE samples.

Compared with the high correlation of VAF between the sequencing platforms used in the fresh-frozen samples ($r = 0.89$, $r^2 = 0.79$, Pearson correlation), the correlation between WES from fresh-frozen samples and HCC panel by using FFPE samples was more modest ($r = 0.67$, $r^2 = 0.45$, Pearson correlation) (Supplemental Figure S2, A and B). Mutations with large deviations in VAFs between the sequencing platforms used in the fresh-frozen samples tended to be covered at reduced depths on either platform (Supplemental Figure S2C). Similar observations could be made between VAFs of exome (fresh-frozen) and HCC panel (FFPE) (Supplemental Figure S2D). The deviations in the latter may be more noticeable by the overall lower depth achieved in the FFPE samples than in the HCC panel sequencing of the fresh-frozen samples. Intratumor heterogeneity between the fresh-frozen and FFPE aliquots likely contributed to the reduced correlation.

Taken together these results suggested that the HCC panel analysis has high specificity and sensitivity in somatic mutation detection. Furthermore, somatic mutations in promoter regions (*TERT* promoter) and lncRNA genes (*MALAT1* and *NEAT1*) highly mutated in HCC could also be detected.

Copy Number Analysis of the HCC Panel Reveals High Concordance with WES

To determine whether the HCC panel could also be used to detect CNAs, 42 genes whose coding regions were entirely covered or were tiled across the lengths of the genes for CNA detection were evaluated (Figure 1A and Supplemental Table S1). Using the 41 nontumoral samples, the variability of the depth of coverage in the amplicons targeting the 42 genes was assessed (*Materials and Methods*). After removing amplicons with low depth of coverage or high variability, 1483 amplicons were used for CNA profiling. To assess the ability to detect per-gene CNA, each nontumoral sample was further paired with two other randomly selected, sex-matched nontumoral samples. The copy number $\log_2$ ratio of five genes, namely *LRP1B*, *ALB*, *BRD7*, *ACVR2A*, and *IRF2*, was variable (SD $>$ 0.3); therefore, these genes were excluded from further CNA analyses. Thirty-seven genes were included in the CNA analysis.

Figure 3 Comparison of somatic mutations defined by whole-exome sequencing (WES) and hepatocellular carcinoma (HCC) panel in fresh-frozen tissues. **A:** Number of coding and noncoding mutations per case identified in 10 fresh-frozen biopsies by using the HCC panel. **B:** Comparison of somatic coding and noncoding mutations found by WES and the HCC panel in the fresh-frozen samples. Heatmaps indicate the variant allele fractions of the somatic mutations (blue, see color key) or their absence (gray) in the eight cases in which at least one somatic mutation was identified. Mutation types are indicated as colored dots according to the color key. Mutations that were not called by mutation caller but were supported by at least one sequencing read are indicated by **asterisks**.

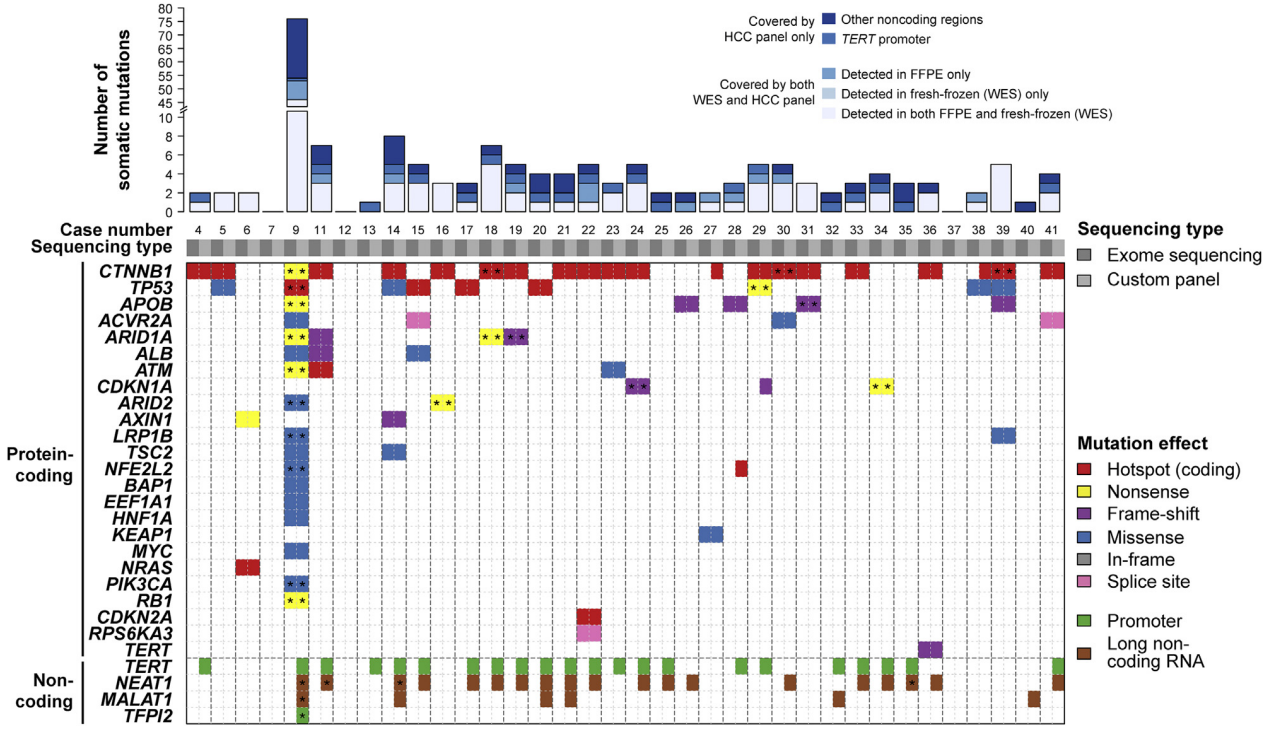


Figure 4 Comparison of somatic mutations defined by whole-exome sequencing (WES) and hepatocellular carcinoma (HCC) panel in formalin-fixed paraffin-embedded (FFPE) tissue. Barplot illustrates the number of somatic coding and noncoding mutations found in 36 FFPE tumor biopsies by using the HCC panel. In the main panel, each row represents a gene on the HCC panel and each column represents a sample. The mutations identified by WES in the fresh-frozen biopsies and those defined by sequencing the corresponding FFPE samples by using the HCC panel are placed next to each other. Mutation types are color coded according to the color key. The presence of multiple mutations in the same gene is illustrated by **asterisks**. Noncoding regions below the **dashed line** were not covered by WES.

The copy number profiles of matched fresh-frozen tumor/nontumor pairs and those derived from WES were compared. Of the 10 fresh-frozen pairs sequenced by using the HCC panel, one was excluded for excessive residual copy number $\log_2$ ratio (segment interquartile range, >0.8).[25] For the nine evaluable samples, a correlation of $r = 0.80$ ($r^2 = 0.64$) was found between the copy number $\log_2$ ratio of the two platforms (Figure 5A). When the copy number profiles of the 34 evaluable FFPE tumors were compared with the matched profiles from WES, a correlation of $r = 0.73$ ($r^2 = 0.54$) was observed between the copy number $\log_2$ ratios (Figure 5A). Overall, 86% of the evaluable genes had concordant copy number states (Figure 5B).

It has previously been reported that tumor purity had an impact on the ability to make CNA calls.[25,34] The impact of tumor purity on CNA analysis was therefore evaluated by using an *in silico* simulation on 12 cases (six fresh-frozen and six FFPE, selected on the basis of the presence of gene amplification/high gain or deep deletion), by replacing tumor reads with reads sampled from the normal samples to simulate tumor content 5%, 10%, 20% up to the actual tumor content for the samples. It was observed that amplifications/high gains were readily detected at 5% tumor content in many cases and at 20% in all cases (Supplemental Figure S5). In this cohort, deep deletions could not be detected at tumor content <40%.

Taken together, these results demonstrated that, despite profiling only a small number of genes, the HCC panel was able to detect CNAs in genes frequently gained or lost in HCC in both fresh-frozen and FFPE tumor samples with low-input DNA.

Discussion

HCC has a distinct mutational landscape compared with the major tumor entities. Numerous genes have been found to be mutated frequently in HCC but rarely in other tumors, such as those important for hepatocyte differentiation (*ALB*, *APOB*, *HNF1A*, *HNF4A*) and inflammatory response (*IL6R*, *IL6ST*). Given the relative rarity of HCC, these genes are currently not targeted or are only partially targeted in commercial panels [eg, Oncomine Comprehensive Panel version 3 (Thermo Fisher Scientific)] and in panels used by sequencing services [eg, FoundationOne assay (Foundation Medicine, Cambridge, MA)] (Supplemental Table S1). Thus, the currently available commercial assays for genomic profiling have suboptimal utility for HCC, and a targeted sequencing panel specifically designed for HCC is warranted.

In this study, we designed a custom Ion Torrent AmpliSeq sequencing panel, targeting all exons of 33 protein-coding genes, two lncRNA genes, promoter regions of four

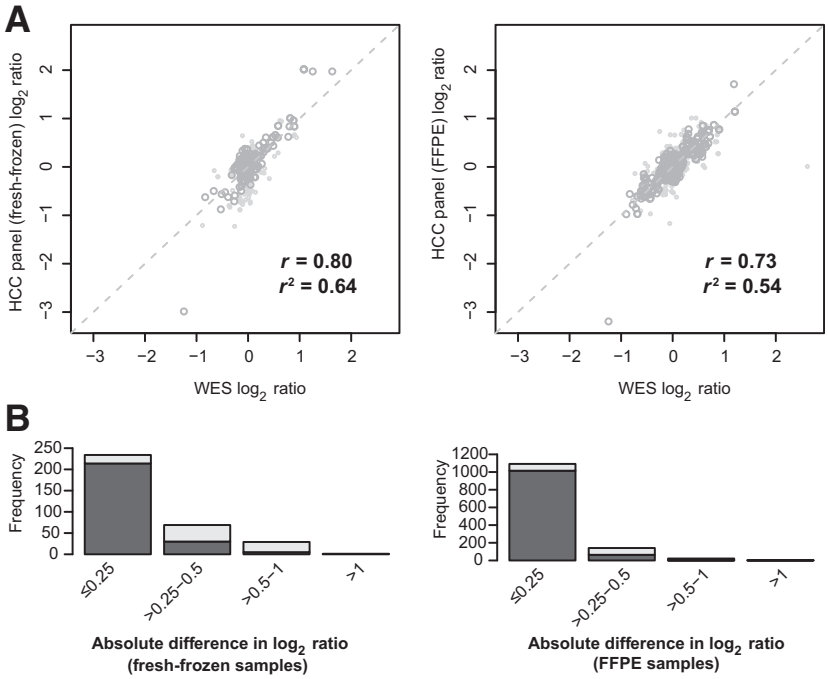


Figure 5 Copy number profiling by using the hepatocellular carcinoma (HCC) panel. **A:** Scatter plots illustrate the copy number $\log_2$ ratio of whole-exome sequencing (WES) and HCC panel sequencing of the fresh-frozen and the formalin-fixed, paraffin-embedded (FFPE) tumor samples. **B:** Barplots illustrate the number of genes with concordant (dark gray) or discordant (light gray) copy number states, binned by the absolute difference in copy number $\log_2$ ratio between WES and HCC panel sequencing of the fresh-frozen and FFPE samples.

genes previously found to be recurrently mutated in HCC, nine genes frequently affected by CNAs, and mutation hotspots in seven cancer genes.[7–17] Of importance, a number of the genes targeted by using the HCC panel are not currently on these two commercial panels. Of the 39 cases profiled with the HCC panel (including both fresh-frozen and FFPE samples), at least one somatic mutation was detected in 90% (35 of 39) of the cases. Of the mutations in coding genes found using this panel, 22% (42 of 189) would have been missed by both Oncomine Comprehensive Panel version 3 and the FoundationOne assay. In addition, recent whole-genome studies of HCC have revealed frequent mutations in lncRNA genes *NEAT1* and *MALAT1*, both of which are not currently targeted by commercial panels. In fact, it was found that approximately one-third of the mutations on the HCC panel were within the promoter and lncRNA regions.

Mutation screening and copy number profiling results from the HCC panel were benchmarked against those obtained from WES by the orthogonal Illumina sequencing technology. All but one mutation identified from WES were detected by using the HCC panel. An additional 10% to 15% of mutations within the coding regions were identified. Most of these additional mutations were in fact supported by few reads by WES; thus, the increased sensitivity was likely a direct result of the increased sequencing depth of both the tumor and the matched normal samples achieved. Crucially, however, evidence of intratumor genetic heterogeneity between adjacent fresh-frozen and FFPE biopsies, including two *CTNNB1* mutations, was found, suggesting that in these cases the *CTNNB1* mutations were not trunk mutations.

Although CNA detection using capture-based methods has been successful for targeted sequencing panel of several hundred genes,[35] CNA detection using amplicon-based targeted sequencing has proven more difficult. A recent study investigated the use of an amplicon-based sequencing strategy that targeted all exons of 113 genes related to DNA repair.[25] The researchers demonstrated that, with an appropriate analysis strategy and quality control, amplicon-based sequencing strategy is feasible and cost-effective for CNA profiling in FFPE samples.[25] In the present study, the strategy of computing and centering the $\log_2$ ratios for the primer two pools separately, before merging and segmentation proved to be an effective strategy in resolving issues associated with variable amplification efficiencies, with 86% of the genes showing concordant copy number states. Considering few studies have investigated the use of small targeted sequencing panel for CNA profiling, further benchmarking studies comparing analysis strategies and including larger sample size will likely improve the accuracies.

In the clinical setting, the quality, type, and amount of input materials for genomic profiling are crucial considerations, particularly in light of the smaller tumors being detected in screening programs. Here, we demonstrated that the HCC panel could be used for genomic screening with high sensitivity and specificity with low-input DNA (20 ng) derived from FFPE samples without compromising the results. Although based on an analysis of the TCGA HCC cases, 92% and 85% of the cases would have exhibited at least one nonsynonymous mutation by using the FoundationOne and the Oncomine assays, respectively, the HCC panel holds the advantage of much lower input requirement

than that required for commercial panels (eg, >40-μm tissue samples for the FoundationOne assay) and for capture-based targeted sequencing strategies.[35] In addition, somatic genetic alterations (somatic mutations and amplifications) could be detected from tumor samples with as low as 30% tumor content. Considering that mutations in the one sample with 5% tumor content could not be detected, 30% may be the lower limit of successful genomic profiling. Although lower limits (approximately 20%) have also been reported,[36] samples were not available to verify this. The samples included in this study are *de facto* samples obtained from routine diagnostic practice, and it was demonstrated that the low-input DNA requirement facilitates genomic profiling from small biopsies.

Driver genetic alterations have not yet become a tangible tool in clinical decision making for the treatment of HCC; thus, the immediate clinical application of our panel may be limited. However, recent studies have described the association of *TERT* promoter and *CTNNB1* exon 3 mutations with increased risk of malignant transformation of hepatocellular adenomas,[37,38] more frequent *HNF1A* and *IL6ST* mutations in hepatocellular adenomas than HCCs,[37] as well as *TP53* mutation as a poor prognostic indicator in HCC.[39–41] These associations suggest a potential utility of genomic profiling in prognostication for hepatocellular adenomas and HCCs, in tissues or even in cell-free DNA.[41,42] In terms of potential targetable alterations, three somatic mutations identified in our cohort of HCC are molecular targets in other cancer types according to OncoKB.[43] These include *ATM* loss of function mutation using olaparib in prostate cancer (level 4; biological evidence), *NRAS* hotspot mutation with binimetinib or in combination with ribociclib in melanoma (level 3; clinical evidence), and *TSC2* mutation with everolimus in central nervous system cancer (level 2; standard of care).[43] Application of our panel in clinical decision may become feasible in the future.

This study has several limitations. First, the targeted nature of the HCC panel means that copy number profiling is not genome-wide and is restricted to the genes included on the panel. Clinically, focal amplifications, compared with gains of chromosome arm, are more likely to be true driver genetic event and may be considered drug targets. The targeted nature of the HCC panel makes it difficult to distinguish the two scenarios. However, a re-analysis of the TCGA data suggests that high-level gains of chr11q13.3 (encompassing *CCND1*, *FGF19*, *FGF3*, *FGF4*) are almost always focal amplifications (>93%), whereas 50% to 70% of high-level gains of *TERT* and *VEGFA* are focal amplifications (Supplemental Table S5). By contrast, high-level gains of chr1q (*SETDB1* and *IL6R*) and chr8q (*NCOA2*, *MYC*, and *PTK2*) are frequently nonfocal (<10%), consistent with the frequent high-level gain of entire arms of chr1q and chr8q.[17] For deletions, most deep deletions are focal deletions, including all deletions (100%) in *ARID2*, *AXIN1*, *CDKN2A/B*, *PTEN*, and *TSC1/2*. These results suggest that CNAs affecting some of the most promising drug targets on the HCC panel are frequently true focal CNAs. Second, given that a median of two to three mutations per tumor were identified, tumor mutational burden, a putative biomarker for response to immune therapy, may not be accurately defined.[44] Third, the HCC panel does not include unique molecular identifiers, which would be useful to assess library complexity, particularly for samples with low-input DNA. We envisage that the addition of unique molecular identifiers would be particularly beneficial for the study of cell-free DNA from HCC patients.[41,42] Fourth, we designed the panel specific for HCC. Recent studies have revealed that mixed HCC/cholangiocarcinoma and cholangiocarcinoma have recurrent mutations in genes such as *IDH1/2*,[45] whereas *FRK* mutations decrease in frequency from hepatocellular adenoma to HCC.[37] These genes are not covered by the HCC panel. However, as an amplicon-based sequencing panel, adding amplicons to include genes that may assist in the differential diagnosis of HCC is straightforward.

Conclusion

This study demonstrated that the HCC panel is a cost-effective strategy for mutation screening and copy number profiling for routine diagnostic HCC samples with low-input DNA.

Acknowledgments

S.P., C.K.Y.N., and L.M.T. conceived and supervised the study; L.Q., M.S.M., S.P., C.K.Y.N., and L.M.T. performed literature search and designed the sequencing panel; S.W. and M.H.H. provided the samples and the whole-exome sequencing data; V.Pa., N.T., M.L., V.Pe., and S.P. performed DNA extraction and sequencing and prepared the library; A.G. and C.K.Y.N. developed the bioinformatics pipeline for mutation calling; V.Pa., A.G., S.P., C.K.Y.N., and L.M.T. analyzed the results and wrote the manuscript.

Supplemental Data

Supplemental material for this article can be found at *https://doi.org/10.1016/j.jmoldx.2018.07.003*.

References

1. Chin L, Andersen JN, Futreal PA: Cancer genomics: from discovery science to personalized medicine. Nat Med 2011, 17:297–303
2. Mok TS, Wu YL, Ahn MJ, Garassino MC, Kim HR, Ramalingam SS, Shepherd FA, He Y, Akamatsu H, Theelen WS, Lee CK, Sebastian M, Templeton A, Mann H, Marotti M, Ghiorghiu S, Papadimitrakopoulou VA; AURA3 Investigators: Osimertinib or platinum-pemetrexed in EGFR T790M-positive lung cancer. N Engl J Med 2017, 376:629–640
3. Toy W, Weir H, Razavi P, Lawson M, Goeppert AU, Mazzola AM, Smith A, Wilson J, Morrow C, Wong WL, De Stanchina E,

Carlson KE, Martin TS, Uddin S, Li Z, Fanning S, Katzenellenbogen JA, Greene G, Baselga J, Chandarlapaty S: Activating ESR1 mutations differentially affect the efficacy of ER antagonists. Cancer Discov 2017, 7:277–287
4. Kris MG, Johnson BE, Berry LD, Kwiatkowski DJ, Iafrate AJ, Wistuba II, Varella-Garcia M, Franklin WA, Aronson SL, Su PF, Shyr Y, Camidge DR, Sequist LV, Glisson BS, Khuri FR, Garon EB, Pao W, Rudin C, Schiller J, Haura EB, Socinski M, Shirai K, Chen H, Giaccone G, Ladanyi M, Kugler K, Minna JD, Bunn PA: Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. JAMA 2014, 311:1998–2006
5. Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, Chandramohan R, Liu ZY, Won HH, Scott SN, Brannon AR, O'Reilly C, Sadowska J, Casanova J, Yannes A, Hechtman JF, Yao J, Song W, Ross DS, Oultache A, Dogan S, Borsu L, Hameed M, Nafa K, Arcila ME, Ladanyi M, Berger MF: Memorial Sloan Kettering-Integrated Mutation Profiling Of Actionable Cancer Targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. J Mol Diagn 2015, 17:251–264
6. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MDM, Miller CA, Welch JS, Walter MJ, Wendl MC, Ley TJ, Wilson RK, Raphael BJ, Ding L: Mutational landscape and significance across 12 major cancer types. Nature 2013, 502:333–339
7. Guichard C, Amaddeo G, Imbeaud S, Ladeiro Y, Pelletier L, Maad IB, Calderaro J, Bioulac-Sage P, Letexier M, Degos F, Clement B, Balabaud C, Chevet E, Laurent A, Couchy G, Letouze E, Calvo F, Zucman-Rossi J: Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat Genet 2012, 44:694–698
8. Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, et al: Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. Nat Genet 2012, 44:760–764
9. Cleary SP, Jeck WR, Zhao X, Chen K, Selitsky SR, Savich GL, Tan TX, Wu MC, Getz G, Lawrence MS, Parker JS, Li J, Powers S, Kim H, Fischer S, Guindi M, Ghanekar A, Chiang DY: Identification of driver genes in hepatocellular carcinoma by exome sequencing. Hepatology 2013, 58:1693–1702
10. Kan Z, Zheng H, Liu X, Li S, Barber TD, Gong Z, et al: Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. Genome Res 2013, 23:1422–1433
11. Ahn SM, Jang SJ, Shim JH, Kim D, Hong SM, Sung CO, Baek D, Haq F, Ansari AA, Lee SY, Chun SM, Choi S, Choi HJ, Kim J, Kim S, Hwang S, Lee YJ, Lee JE, Jung WR, Jang HY, Yang E, Sung WK, Lee NP, Mao M, Lee C, Zucman-Rossi J, Yu E, Lee HC, Kong G: Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. Hepatology 2014, 60:1972–1982
12. Jhunjhunwala S, Jiang Z, Stawiski EW, Gnad F, Liu J, Mayba O, Du P, Diao J, Johnson S, Wong KF, Gao Z, Li Y, Wu TD, Kapadia SB, Modrusan Z, French DM, Luk JM, Seshagiri S, Zhang Z: Diverse modes of genomic alteration in hepatocellular carcinoma. Genome Biol 2014, 15:436
13. Totoki Y, Tatsuno K, Covington KR, Ueda H, Creighton CJ, Kato M, et al: Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nat Genet 2014, 46:1267–1273
14. Shiraishi Y, Fujimoto A, Furuta M, Tanaka H, Chiba K, Boroevich KA, Abe T, Kawakami Y, Ueno M, Gotoh K, Ariizumi S, Shibuya T, Nakano K, Sasaki A, Maejima K, Kitada R, Hayami S, Shigekawa Y, Marubashi S, Yamada T, Kubo M, Ishikawa O, Aikata H, Arihiro K, Ohdan H, Yamamoto M, Yamaue H, Chayama K, Tsunoda T, Miyano S, Nakagawa H: Integrated analysis of whole genome and transcriptome sequencing reveals diverse transcriptomic aberrations driven by somatic genomic changes in liver cancers. PLoS One 2014, 9:e114263
15. Schulze K, Imbeaud S, Letouze E, Alexandrov LB, Calderaro J, Rebouissou S, Couchy G, Meiller C, Shinde J, Soysouvanh F, Calatayud AL, Pinyol R, Pelletier L, Balabaud C, Laurent A, Blanc JF, Mazzaferro V, Calvo F, Villanueva A, Nault JC, Bioulac-Sage P, Stratton MR, Llovet JM, Zucman-Rossi J: Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. Nat Genet 2015, 47:505–511
16. Fujimoto A, Furuta M, Totoki Y, Tsunoda T, Kato M, Shiraishi Y, et al: Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. Nat Genet 2016, 48:500–509
17. Cancer Genome Atlas Research Network: Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell 2017, 169:1327–1341.e23
18. Weinhold N, Jacobsen A, Schultz N, Sander C, Lee W: Genome-wide analysis of noncoding regulatory mutations in cancer. Nat Genet 2014, 46:1160–1165
19. Ng CK, Piscuoglio S, Geyer FC, Burke KA, Pareja F, Eberle C, Lim R, Natrajan R, Riaz N, Mariani O, Norton L, Vincent-Salomon A, Wen YH, Weigelt B, Reis-Filho JS: The landscape of somatic genetic alterations in metaplastic breast carcinomas. Clin Cancer Res 2017, 23:3859–3870
20. Piscuoglio S, Ng CK, Murray MP, Guerini-Rocco E, Martelotto LG, Geyer FC, Bidard FC, Berman S, Fusco N, Sakr RA, Eberle CA, De Mattos-Arruda L, Macedo GS, Akram M, Baslan T, Hicks JB, King TA, Brogi E, Norton L, Weigelt B, Hudis CA, Reis-Filho JS: The genomic landscape of male breast cancers. Clin Cancer Res 2016, 22:4045–4056
21. Chang MT, Asthana S, Gao SP, Lee BH, Chapman JS, Kandoth C, Gao J, Socci ND, Solit DB, Olshen AB, Schultz N, Taylor BS: Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. Nat Biotechnol 2016, 34:155–163
22. Gao J, Chang MT, Johnsen HC, Gao SP, Sylvester BE, Sumer SO, Zhang H, Solit DB, Taylor BS, Schultz N, Sander C: 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. Genome Med 2017, 9:4
23. Thorvaldsdottir H, Robinson JT, Mesirov JP: Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2013, 14:178–192
24. Talevich E, Shain AH, Botton T, Bastian BC: CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput Biol 2016, 12:e1004873
25. Seed G, Yuan W, Mateo J, Carreira S, Bertan C, Lambros M, Boysen G, Ferraldeschi R, Miranda S, Figueiredo I, Riisnaes R, Crespo M, Rodrigues DN, Talevich E, Robinson DR, Kunju LP, Wu YM, Lonigro R, Sandhu S, Chinnayan A, de Bono JS: Gene copy number estimation from targeted next-generation sequencing of prostate cancer biopsies: analytic validation and clinical qualification. Clin Cancer Res 2017, 23:6070–6077
26. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK: VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012, 22:568–576
27. Olshen AB, Venkatraman ES, Lucito R, Wigler M: Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 2004, 5:557–572
28. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009, 25:1754–1760
29. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA: The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010, 20:1297–1303
30. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G: Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013, 31:213–219

31. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK: Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 2012, 28: 1811−1817
32. Shen R, Seshan VE: FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res 2016, 44:e131
33. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013, 6:pl1
34. Grasso C, Butler T, Rhodes K, Quist M, Neff TL, Moore S, Tomlins SA, Reinig E, Beadling C, Andersen M, Corless CL: Assessing copy number alterations in targeted, amplicon-based next-generation sequencing data. J Mol Diagn 2015, 17:53−63
35. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al: Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med 2017, 23:703−713
36. Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, et al: Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol 2013, 31:1023−1031
37. Pilati C, Letouze E, Nault JC, Imbeaud S, Boulai A, Calderaro J, Poussin K, Franconi A, Couchy G, Morcrette G, Mallet M, Taouji S, Balabaud C, Terris B, Canal F, Paradis V, Scoazec JY, de Muret A, Guettier C, Bioulac-Sage P, Chevet E, Calvo F, Zucman-Rossi J: Genomic profiling of hepatocellular adenomas reveals recurrent FRK-activating mutations and the mechanisms of malignant transformation. Cancer Cell 2014, 25:428−441
38. Nault JC, Couchy G, Balabaud C, Morcrette G, Caruso S, Blanc JF, Bacq Y, Calderaro J, Paradis V, Ramos J, Scoazec JY, Gnemmi V, Sturm N, Guettier C, Fabre M, Savier E, Chiche L, Labrune P, Selves J, Wendum D, Pilati C, Laurent A, De Muret A, Le Bail B, Rebouissou S, Imbeaud S; GENTHEP Investigators, Bioulac-Sage P, Letouze E, Zucman-Rossi J: Molecular classification of hepatocellular adenoma associates with risk factors, bleeding, and malignant transformation. Gastroenterology 2017, 152:880−894.e6
39. Goossens N, Sun X, Hoshida Y: Molecular classification of hepatocellular carcinoma: potential therapeutic implications. Hepat Oncol 2015, 2:371−379
40. Desert R, Rohart F, Canal F, Sicard M, Desille M, Renaud S, Turlin B, Bellaud P, Perret C, Clement B, Le Cao KA, Musso O: Human hepatocellular carcinomas with a periportal phenotype have the lowest potential for early recurrence after curative resection. Hepatology 2017, 66:1502−1518
41. Kancherla V, Abdullazade S, Matter MS, Lanzafame M, Quagliata L, Roma G, Hoshida Y, Terracciano LM, Ng CKY, Piscuoglio S: Genomic analysis revealed new oncogenic signatures in TP53-mutant hepatocellular carcinoma. Front Genet 2018, 9:2
42. Ng CKY, Di Costanzo GG, Tosti N, Paradiso V, Coto-Llerena M, Roscigno G, Perrina V, Quintavalle C, Boldanova T, Wieland S, Marino-Marsilia G, Lanzafame M, Quagliata L, Condorelli G, Matter MS, Tortora R, Heim MH, Terracciano LM, Piscuoglio S: Genetic profiling using plasma-derived cell-free DNA in therapy-naive hepatocellular carcinoma patients: a pilot study. Ann Oncol 2018, 29:1286−1291
43. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, et al: OncoKB: a precision oncology knowledge base. JCO Precis Oncol 2017, 2017
44. Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, Stephens PJ, Daniels GA, Kurzrock R: Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. Mol Cancer Ther 2017, 16:2598−2608
45. Farshidfar F, Zheng S, Gingras MC, Newton Y, Shih J, Robertson AG, et al: Integrative genomic analysis of cholangiocarcinoma identifies distinct IDH-mutant molecular profiles. Cell Rep 2017, 19:2878−2880

9.2 Genetic profiling using plasma-derived cell-free DNA in therapy-naïve hepatocellular carcinoma patients: a pilot study.

Annals of Oncology 29: 1286–1291, 2018
doi:10.1093/annonc/mdy083
Published online 2 March 2018

ORIGINAL ARTICLE

Genetic profiling using plasma-derived cell-free DNA in therapy-naïve hepatocellular carcinoma patients: a pilot study

C. K. Y. Ng[1,2*], G. G. Di Costanzo[3], N. Tosti[1], V. Paradiso[1], M. Coto-Llerena[2], G. Roscigno[4], V. Perrina[1], C. Quintavalle[1], T. Boldanova[2,5], S. Wieland[2], G. Marino-Marsilia[6], M. Lanzafame[1], L. Quagliata[1], G. Condorelli[4], M. S. Matter[1], R. Tortora[3], M. H. Heim[2,5], L. M. Terracciano[1] & S. Piscuoglio[1*]

[1]Institute of Pathology, University Hospital Basel, Basel; [2]Hepatology Laboratory, Department of Biomedicine, University of Basel, Basel, Switzerland; [3]Department of Transplantation – Liver Unit, Cardarelli Hospital, Naples; [4]Department of Molecular Medicine and Medical Biotechnology, "Federico II" University of Naples, Naples, Italy; [5]Division of Gastroenterology and Hepatology, University Hospital Basel, Basel, Switzerland; [6]Pathology Unit, Cardarelli Hospital, Naples, Italy

Correspondence to: Dr Charlotte K. Y. Ng, Institute of Pathology, University Hospital Basel, Schoenbeinstrasse 40, 4031 Basel, Switzerland. Tel: +41-613286874; Fax: +41-612653194; E-mail: kiuyancharlotte.ng@usb.ch

Dr Salvatore Piscuoglio, Institute of Pathology, University Hospital Basel, Schoenbeinstrasse 40, 4031 Basel, Switzerland. Tel: +41-613286874; Fax: +41-612653194; E-mail: salvatore.piscuoglio@usb.ch

Background: Hepatocellular carcinomas (HCCs) are not routinely biopsied, resulting in a lack of tumor materials for molecular profiling. Here we sought to determine whether plasma-derived cell-free DNA (cfDNA) captures the genetic alterations of HCC in patients who have not undergone systemic therapy.

Patients and methods: Frozen biopsies from the primary tumor and plasma were synchronously collected from 30 prospectively recruited, systemic treatment-naïve HCC patients. Deep sequencing of the DNA from the biopsies, plasma-derived cfDNA and matched germline was carried out using a panel targeting 46 coding and non-coding genes frequently altered in HCCs.

Results: In 26/30 patients, at least one somatic mutation was detected in biopsy and/or cfDNA. Somatic mutations in HCC-associated genes were present in the cfDNA of 63% (19/30) of the patients and could be detected 'de novo' without prior knowledge of the mutations present in the biopsy in 27% (8/30) of the patients. Mutational load and the variant allele fraction of the mutations detected in the cfDNA positively correlated with tumor size and Edmondson grade. Crucially, among the seven patients in whom the largest tumor was ≥5 cm or was associated with metastasis, at least one mutation was detected 'de novo' in the cfDNA of 86% (6/7) of the cases. In these patients, cfDNA and tumor DNA captured 87% (80/92) and 95% (87/92) of the mutations, suggesting that cfDNA and tumor DNA captured similar proportions of somatic mutations.

Conclusion: In patients with high disease burden, the use of cfDNA for genetic profiling when biopsy is unavailable may be feasible. Our results support further investigations into the clinical utility of cfDNA in a larger cohort of patients.

Key words: hepatocellular carcinoma, cell-free DNA, circulating tumor DNA, somatic mutations, liquid biopsy, mutation screening

Introduction

The invasive nature of biopsy has prompted investigations into the use of plasma-derived cell-free DNA (cfDNA) as a potential minimally invasive surrogate for molecular profiling in several cancer types [1–4]. In contrast to most solid tumor types, hepatocellular carcinoma (HCC) diagnosis is frequently on the basis of radiology alone and in the absence of tumor biopsy. Therefore, nucleic acids for genetic profiling of HCC are typically obtained from tumor resection, a procedure that is only carried out in patients with limited, early-stage disease. In unresectable HCC patients, should the need for molecular profiling arises, the tumor materials would have to be collected in a non-routine invasive

procedure. The lack of routinely collected tumor materials is a hurdle for wider adoption of tumor profiling.

Studies have found that cfDNA concentration in serum or plasma of HCC patients is 3–4 times higher than in patients with chronic hepatitis and is up to 20 times higher than in healthy individuals [5–7]. Moreover, cfDNA concentration was found to be associated with tumor size, portal vein invasion and may be prognostic [5–8]. Molecular studies of circulating tumor DNA (ctDNA) in HCCs have investigated the size profiles of ctDNA [9], or were mutational studies of few cases, of resected materials, carried out at very low depth or investigated few mutation hotspots [10–14]. The use of resected materials, however, restricts molecular analyses to patients with early-stage, resectable disease. Given the correlation of cfDNA concentration and tumor size, one may speculate that patients with later stage disease would have higher mutational burden in cfDNA, as has been shown in other cancer types [4, 15].

Restricting molecular studies of ctDNA to mutation hotspots risks missing a substantial number of mutations, as most somatic mutations in HCC, even those in HCC-associated driver genes, do not fall into mutation hotspots [16–21]. Besides *TP53* (p53), *CTNNB1* (β-catenin) and *TERT* promoter, a wide range of HCC-associated driver genes and recurrently mutated promoter regions have been discovered, including those involved in chromatin remodeling (e.g. *ARID1A*, *ARID1B*, *ARID2*, *BAP1*), Wnt/β-catenin pathway (e.g. *AXIN1, FGF19*), and response to oxidative stress (e.g. *KEAP1*, *NFE2L2*) [16–21]. Additionally, long non-coding RNA genes (lncRNA, e.g. *NEAT1*, *MALAT1*) and promoter regions of *WDR74, TFPI2 and MED16* are also recurrently mutated [18, 19, 22].

In this exploratory study, we sought to determine whether somatic mutations in HCC driver genes can be detected with high confidence using next-generation sequencing in the plasma-derived cfDNA of HCC patients who have not undergone systemic therapy, and if the repertoire of mutations in the cfDNA is representative of the synchronously collected tumor biopsy. To address these questions, we prospectively recruited 30 HCC patients from whom we synchronously collected diagnostic core needle tumor biopsy and whole blood (supplementary Table S1, available at *Annals of Oncology* online) and carried out deep sequencing targeting HCC driver genes and mutation hotspots (supplementary Table S2, available at *Annals of Oncology* online).

Patients and methods

Patients

Thirty patients diagnosed with HCC at the University Hospital Basel, Basel, Switzerland or at Ospedale Cardarelli, Naples, Italy, were prospectively recruited for this study after written informed consent (supplementary Table S1, available at *Annals of Oncology* online). Patients who had previous systemic therapy for HCC were excluded. One patient was treated with radio-frequency thermal ablation 21 months before sample collection. From each patient undergoing diagnostic liver biopsy, two ultrasound-guided core needle biopsies of the primary tumor and whole blood were collected at diagnosis at the same time. Of the two primary tumor biopsies, one was processed and embedded in paraffin for clinical purposes and the other one was snap-frozen and stored at −80°C for research purposes. Ten millilitres of whole blood was collected in a 10 ml Cell-Free DNA Blood Collection Tube (BCT, Streck) and processed immediately (supplementary methods, available at *Annals of Oncology* online). Plasma was stored at −80°C until cfDNA extraction.

Tumor size, tumor location, macrovascular invasion, multifocality, and extrahepatic spread of each patient were assessed radiologically. Clinical staging of the patients was determined according to the Barcelona Clinic Liver Cancer (BCLC) staging system [23]. Sex of the patients, serum alpha-fetoprotein (AFP) levels, primary risk factors (hepatitis B/C virus infection, alcoholic liver disease, non-alcoholic fatty liver disease) were retrieved from clinical files. Histologic grading was carried out according to the 4-point scale Edmondson and Steiner system [24] (supplementary methods, available at *Annals of Oncology* online). Approval for the use of these samples has been granted by the ethics committee (Protocol Number EKNZ 2014-099).

Targeted sequencing and analysis

Tumor and germline DNA was extracted from fresh frozen biopsies and peripheral blood leukocytes ('buffy coat'). Circulating cfDNA was extracted from 3 to 6 ml of plasma (supplementary methods, available at *Annals of Oncology* online). DNA samples from the tumors, plasma-derived cfDNA and germline DNA were subjected to targeted sequencing using an Ampliseq panel targeting all exons of 33 liver cancer-associated protein-coding genes, all exons of the recurrently mutated lncRNA genes *MALAT1* and *NEAT1*, recurrently mutated promoter region of *TERT, WDR74, TFPI2* and *MED16*, as well as hotspots mutations in an additional seven cancer genes (supplementary Table S2, available at *Annals of Oncology* online). Sequencing was carried out on an Ion 530 chip using the Ion S5 XL system (Thermo Fisher Scientific, supplementary methods and Table S3, available at *Annals of Oncology* online). Sequencing data have been deposited in the Sequence Read Archive under the accession SRP115181.

Sequence reads were aligned to the human reference genome hg19 using TMAP. Somatic mutations were defined using Torrent Variant Caller (TVC) v5.0.3. We filtered out mutations supported by $\leq$8 reads, and/or those covered by $<$10 reads in the tumor/cfDNA or $<$10 reads in the matched germline. We only retained mutations for which the tumor variant allele fraction (VAF) was at least 10 times that of the matched normal VAF to ensure we kept only the somatic variants (supplementary methods, available at *Annals of Oncology* online). Due to the repetitive nature and the high GC content of the *TERT* promoter region, *TERT* mutation hotspots (chr5: 1295228 and chr5: 1295250) were additionally screened, and were considered present if supported by at least 5 reads or VAF of at least 5%. Mutations identified using the above steps are referred to as those found by 'de novo' methods.

To account for somatic mutations that may be present at low VAF in either the tumor biopsy or the matched cfDNA samples but not both, all somatic mutations identified using the 'de novo' methods in one of the two samples were interrogated for their presence in the matched sample by supplying TVC with their positions as the 'hotspot list'. Mutations supported by at least 2 reads were considered to be present. Mutations identified using the above steps are referred to as those found by 'interrogation'. Clinical actionability was assessed using OncoKB [25].

Statistical analysis

All statistical analyses were carried out in R v3.3.1. Correlations between the number of mutations, cfDNA concentrations and continuous/ordinal clinical variables (supplementary methods, available at *Annals of Oncology* online) were assessed using the Spearman's ρ. Comparisons of continuous/ordinal clinical variables between patients with and without somatic mutations in the cfDNA were carried out using Mann–Whitney *U* tests. Comparisons of categorical clinical variables and between patients with and without somatic mutations in the cfDNA were carried out using Fisher's exact tests. All statistical tests were two-tailed and $P < 0.05$ was considered statistically significant; 95% confidence intervals (CIs)

were estimated by leaving out 20% of the data points, computed over 100 runs.

Results

Of the 30 patients prospectively recruited into this study, 33% (10/30) had BCLC stages B/C/D disease (Table 1; supplementary Table S1, available at *Annals of Oncology* online). Multifocal and metastatic diseases were seen in 11 and 1 patients, respectively. Median diameter of the largest tumor was 34 mm (range 13–220 mm). At least one primary risk factor was identified for all patients (except HPU025 for whom the information is unavailable). Cirrhosis was seen in 87% (26/30) of the cases.

From each patient undergoing diagnostic liver biopsy, a core needle biopsy and whole blood were collected at the same time for targeted sequencing. A median of 94.6 ng (range 19.8–1710 ng) of plasma cfDNA was obtained from 10 ml of whole blood per patient (supplementary Table S1, available at *Annals of Oncology* online). We carried out deep sequencing of the HCC biopsies, cfDNA and matched germline using an in-house custom-made panel targeting 46 coding and lncRNA genes frequently altered in HCCs (median 1339× in biopsies and plasma, range 703–9385×, supplementary Tables S2 and S3, available at *Annals of Oncology* online). To mimic the potential use of plasma-derived cfDNA in the absence of available resected tumor material or a core needle biopsy in a clinical setting, we defined the somatic mutations for each HCC and cfDNA samples independently without prior knowledge of the repertoire of mutations present in the biopsy/cfDNA counterpart following a stringent set of analysis criteria (or 'de novo'). Additionally, to account for mutations that may be present at frequencies below the detection limit of the 'de novo' approach, we further examined the sequencing data of the biopsies for all mutations detected in the cfDNA (or 'by interrogation'), and *vice versa*. In 26/30 patients, at least one somatic mutation was detected in the biopsy and/or cfDNA (Figure 1 and supplementary Table S4, available at *Annals of Oncology* online).

Table 1. Clinicopathologic parameters of 30 therapy-naïve HCC included in this study

Age (*N* = 30)	Median years	72 (49–86)
Gender (*N*=30)	Female	10
	Male	20
BCLC classification (*N*=30)	A	20
	B	8
	C	1
	D	1
Associated with cirrhosis (*N*=30)	Yes	26
	No	4
Edmondson grade (*N*=28)	2	17
	3	7
	4	4
Largest tumor diameter (mm) (*N*=30)	Median (mm)	34 (13–220)
Macrovascular invasion (*N*=29)	Absent	28
	Present	1
Presence of metastasis (*N*=30)	Absent	29
	Present	1
Multifocal (*N*=30)	Absent	19
	Present	11
AFP (ng/ml) (*N*=29)	Median (ng/ml)	9 (1.6–7852)
Macrovascular invasion (*N*=29)	Absent	28
	Present	1
HBV (*N*=29)	Absent	27
	Present	2
HCV (*N*=29)	Absent	12
	Present	17
ALD (*N*=29)	Absent	19
	Present	10
NAFLD (*N*=29)	Absent	26
	Present	3
Prior treatment (*N*=30)	No	29
	Yes	1 (RFTA)

AFP, alpha-fetoprotein; ALD, alcoholic liver disease; BCLC, Barcelona Clinic Liver Cancer; HBV, hepatitis B virus; HCV, hepatitis C virus; NAFLD, non-alcoholic fatty liver disease; RFTA, radio-frequency thermal ablation.

Using the 'de novo' approach, we detected at least one somatic mutation in the cfDNA of 27% (8/30) of the patients (median 3, blue/gold bars, Figure 1). Considering the 7 non-hypermutator cases with at least one detectable mutation in the cfDNA, 81% (17/21) of the mutations detected in the cfDNA were also independently detected in the biopsy counterparts. In the hypermutator case (HPU207), 97% (64/66) of the mutations detected in the cfDNA were also independently detected in the biopsy counterpart (blue bars, Figure 1). All six apparently cfDNA-specific mutations were found to be present at low frequencies in their biopsy counterparts by interrogation (gold bars, Figure 1), suggesting that, in accordance with a recent study [13], cfDNA may be useful in overcoming intra-tumor genetic heterogeneity within the biopsies in therapy-naïve HCC patients. On the other hand, of all mutations detected in the non-hypermutator and the hypermutator cases, 78% (78/100) and 7% (5/71), respectively, were detected only in the HCC biopsies using the 'de novo' approach (dark/light red bars, Figure 1). However, 31% (24/78) and 100% (5/5) of these mutations could in fact be detected in the cfDNA by interrogation (dark red bars, Figure 1). Taken together, these results demonstrate that at least one somatic mutation can be detected in the cfDNA without prior knowledge of the repertoire of mutations in the HCC biopsies in 27% (8/30) of HCC patients and that at least one mutation was present, including those identified by interrogation, in 63% (19/30) of the cases.

Comparing the clinicopathologic parameters, we found that the 8 cases for whom at least one somatic mutation was detected in the cfDNA using the 'de novo' approach were associated with larger tumors (diameter of the largest tumor) and increasing Edmondson grade ($P = 0.012$ and $P = 0.010$, Mann–Whitney U tests, Figure 1; supplementary Table S5, available at *Annals of Oncology* online). Across all patients, the number of mutations detected 'de novo' in the cfDNA was positively correlated with the diameter of the largest tumor and Edmondson grade ($r = 0.482$, $P = 0.007$ and $r = 0.470$, $P = 0.012$, respectively, Spearman's ρ). The diameter of the largest tumor and Edmondson grade were also correlated with the maximum variant allele fractions of the mutations detected in the cfDNA ($r = 0.496$, $P = 0.005$ and $r = 0.502$, $P = 0.007$, respectively,

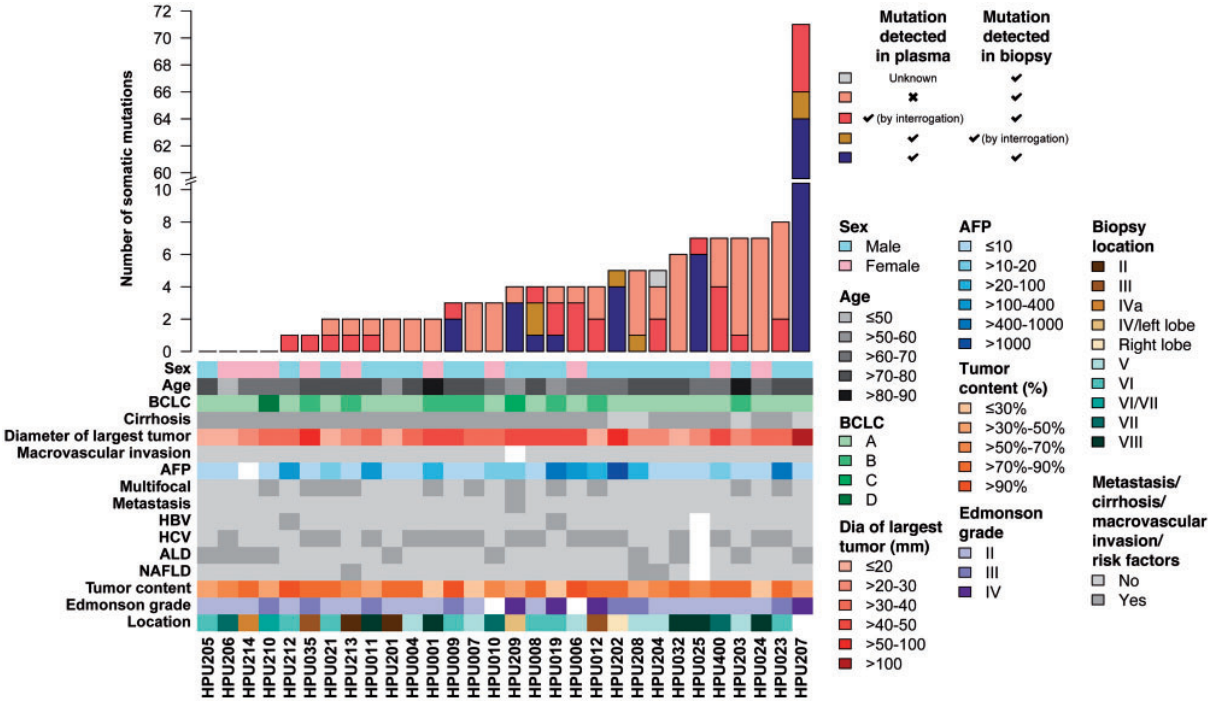


Figure 1. Number of somatic mutations detected in plasma-derived cell-free DNA and clinicopathologic information of the 30 patients with therapy-naïve hepatocellular carcinoma. The number of somatic mutations were categorized based on whether they were detected 'de novo' or 'by interrogation' (i.e. without or with prior knowledge of the repertoire of mutations in the biopsy/cfDNA counterpart, respectively, see color key). Clinicopathologic information is color-coded. White indicates unavailable information.

Spearman's ρ) and cfDNA concentration ($r = 0.889$, $P < 0.001$ and $r = 0.439$, $P = 0.020$, respectively, Spearman's ρ; supplementary Table S5, available at *Annals of Oncology* online). Additionally, at least one mutation was detected 'de novo' in the cfDNA in 40% (8/20) of male patients compared with 0% (0/10) of female patients, and in 75% (3/4) of HCCs not associated with cirrhosis compared with 19% (5/26) of HCCs with cirrhosis ($P = 0.029$ and $P = 0.048$, respectively, Fisher's exact tests; supplementary Table S5, available at *Annals of Oncology* online).

Among the seven cases in whom the largest tumor was ≥5 cm or was associated with metastasis, at least one mutation was detected 'de novo' in the cfDNA of 86% (6/7) of the cases, with a median of 75% (range 0%–100%) of the mutations detected in the cfDNA (Figure 2). Importantly, 87% (80/92, 95% CI 84% to 91%) of the mutations were detected 'de novo', and all but two remaining mutations could be detected by interrogation in the cfDNA counterparts. Conversely, 95% (87/92, 95% CI 93% to 97%) of the mutations were detected 'de novo' in the tumor biopsies, suggesting that mutation profiling of cfDNA in these patients captured similar proportion of mutations as tumor profiling would. By contrast, only 9% (7/78, 95% CI 6% to 11%) of the mutations were detected in the cfDNA of the remaining 23 patients with small (largest tumor ≤5 cm), non-metastatic HCC. These results suggest that in most HCC patients with high tumor burden, somatic mutations can be detected in the cfDNA with high confidence and that the repertoire of somatic mutations detected in cfDNA is representative of that in the primary HCC biopsy.

Discussion

HCC differs from most other tumor types in that biopsies are rarely carried out as they are usually not required for diagnosis. Thus, in patients not eligible for tumor resection (i.e. patients with large or metastatic disease and/or with poor performance status), tumor materials are usually unavailable for molecular profiling. Here we describe a prospective study to investigate the utility of cfDNA collected at the time of biopsy for molecular profiling in HCC patients. Targeting the most significantly mutated genes and regions in HCCs, we found that, even without the prior knowledge of the somatic mutations in the HCCs, high-depth sequencing analysis of plasma-derived cfDNA revealed that at least one somatic mutation in HCC driver genes can be detected in 27% (8/30) of therapy-naïve HCC patients. In an additional 11 cases, cfDNA captured mutations present below 'de novo' detection limit in the biopsies, demonstrating that somatic mutations were present in the cfDNA of 63% (19/30) of HCC patients at diagnosis. Importantly, among the patients with high disease burden (large tumor or metastasis) and most likely to be ineligible for resection, cfDNA profiling captured nearly as many mutations as primary tumor biopsy profiling alone. Of note, a *TSC2* frameshift mutation detected in the cfDNA and the primary tumor of the metastatic patient HPU209 is targetable by everolimus in cancers of the central nervous system as standard of care (supplementary Table S4, available at *Annals of Oncology* online). Taken together, our results demonstrate that the repertoire of mutations in HCC-associated genes identified in the cfDNA is representative of that in the biopsy.

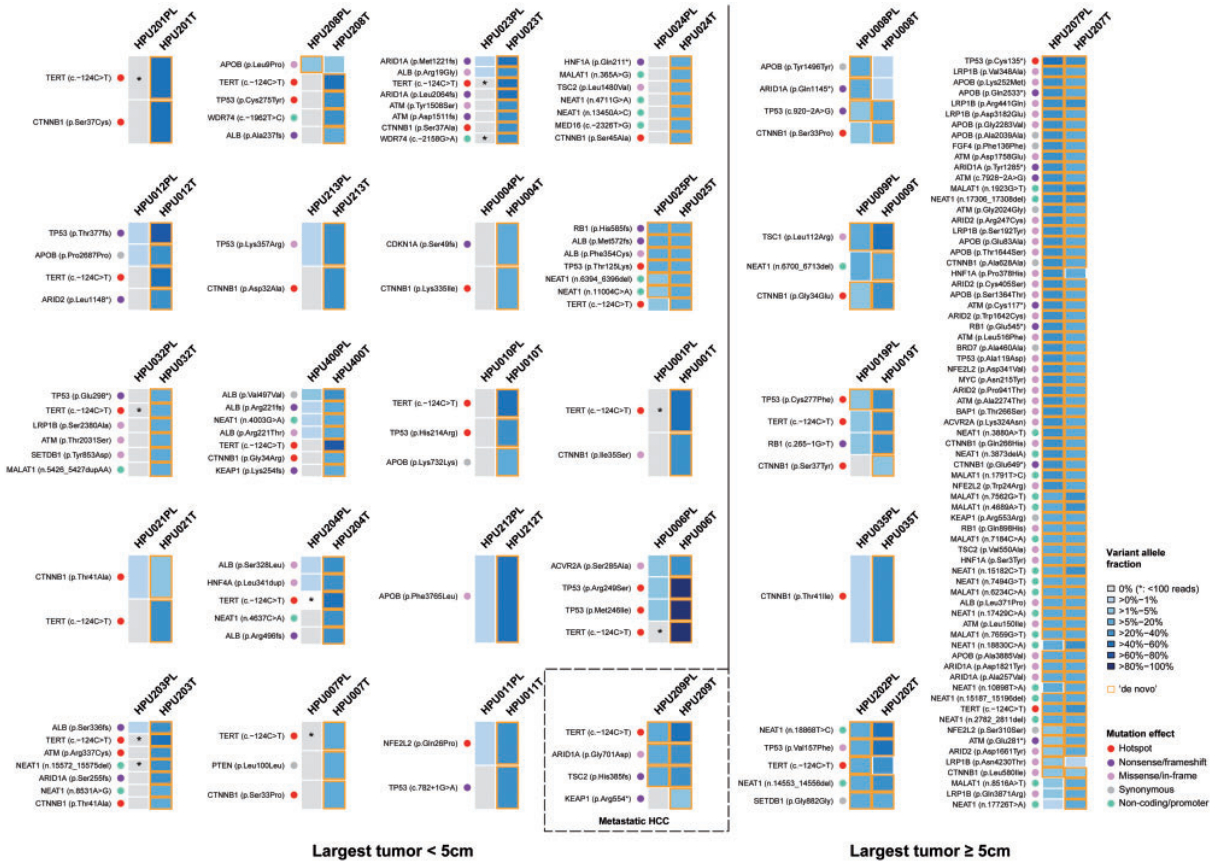


Figure 2. Somatic mutations found in cfDNA and in their primary tumor biopsies. Heatmaps indicate the variant allele fractions of the somatic mutations (blue, see color key) or their absence (grey) in the 26 pairs of tumor biopsy and cfDNA for which at least one somatic mutation was identified. Mutation types are indicated as colored dots. Orange boxes denote the mutations detected using the 'de novo' approach (i.e. without prior knowledge of the mutations in the biopsy/cfDNA counterpart). Mutations not detected by the 'de novo' approach but were covered by <100 reads are indicated by an asterisk. Cases are grouped according to the diameter of the largest tumor. T, tumor; PL, plasma.

Many HCC patients present with multifocal or metastatic disease and variable levels of heterogeneity with branched and parallel evolutionary patterns have been detected in HCC patients [13, 14]. Here we found a number of mutations that were detected with high confidence in the cfDNA but could only be detected by interrogation in the biopsy counterparts, reinforcing the notion that genetic analysis of a single diagnostic biopsy of the primary tumor may not be representative of the disease. Studies into the use of cfDNA as a minimally invasive surrogate for molecular profiling in HCC patients are therefore of particular clinical relevance.

Our study was limited in cohort size but as a proof of principle study and interpreted in the context of other tumor types [1–3], we found strong evidence that somatic mutations can be reliably detected in patients with high disease burden. As a prospective study, we have not assessed the prognostic significance of our findings. Furthermore, our filtering steps for the 'de novo' approach were deliberately stringent to closely recapitulate a potential clinical scenario. It is plausible that the limited sensitivity in detecting mutations 'de novo' in patients with low tumor burden is related to stringent filters. In fact, the number of mutations detected by interrogation suggests that advanced sequencing technologies incorporating molecular barcoding or alternative high-fidelity sequencing techniques will likely increase detection sensitivity in the clinical setting. Despite these limitations, the observed correlation of detectable somatic mutations and disease burden has important implications in the implementation of precision medicine [3]. Our results point toward the use of cfDNA for genetic profiling in HCC patients ineligible for resection and provide an argument for not subjecting patients with high disease burden to otherwise diagnostically unnecessary invasive procedure. Our results support further investigations into the clinical utility of cfDNA in a larger cohort of patients.

Funding

This work was supported by the Krebsliga beider Basel (KLbB-4183-03-2017 to CKYN). Additional financial support was provided by the Swiss Cancer League (Oncosuisse) (KLS-3639-02-2015 to LMT and KFS-3995-08-2016 to SP); the Swiss National Science Foundation (Ambizione PZ00P3_168165 to SP); the European Research Council (ERC Synergy Grant 609883 to MHH).

Disclosure

The authors have declared no conflicts of interest.

References

1. Dawson SJ, Tsui DW, Murtaza M et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med 2013; 368(13): 1199–1209.
2. Siravegna G, Mussolin B, Buscarino M et al. Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. Nat Med 2015; 21(7): 795–801.
3. Bidard FC, Weigelt B, Reis-Filho JS. Going with the flow: from circulating tumor cells to DNA. Sci Transl Med 2013; 5(207): 207ps14.
4. Bettegowda C, Sausen M, Leary RJ et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med 2014; 6: 224ra224.
5. Huang Z, Hua D, Hu Y et al. Quantitation of plasma circulating DNA using quantitative PCR for the detection of hepatocellular carcinoma. Pathol Oncol Res 2012; 18(2): 271–276.
6. Tokuhisa Y, Iizuka N, Sakaida I et al. Circulating cell-free DNA as a predictive marker for distant metastasis of hepatitis C virus-related hepatocellular carcinoma. Br J Cancer 2007; 97(10): 1399–1403.
7. Yang YJ, Chen H, Huang P et al. Quantification of plasma *hTERT* DNA in hepatocellular carcinoma patients by quantitative fluorescent polymerase chain reaction. CIM 2011; 34(4): 238.
8. Iizuka N, Sakaida I, Moribe T et al. Elevated levels of circulating cell-free DNA in the blood of patients with hepatitis C virus-associated hepatocellular carcinoma. Anticancer Res 2006; 26: 4713–4719.
9. Jiang P, Chan CW, Chan KC et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. Proc Natl Acad Sci USA 2015; 112(11): E1317–E1325.
10. Chan KC, Jiang P, Zheng YW et al. Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. Clin Chem 2013; 59(1): 211–224.
11. Kirk GD, Lesi OA, Mendy M et al. 249(ser) *TP53* mutation in plasma DNA, hepatitis B viral infection, and risk of hepatocellular carcinoma. Oncogene 2005; 24(38): 5858–5867.
12. Liao W, Yang H, Xu H et al. Noninvasive detection of tumor-associated mutations from circulating cell-free DNA in hepatocellular carcinoma patients by targeted deep sequencing. Oncotarget 2016; 7(26): 40481–40490.
13. Huang A, Zhao X, Yang XR et al. Circumventing intratumoral heterogeneity to identify potential therapeutic targets in hepatocellular carcinoma. J Hepatol 2017; 67(2): 293–301.
14. Huang A, Zhang X, Zhou SL et al. Detecting circulating tumor DNA in hepatocellular carcinoma patients using droplet digital PCR is feasible and reflects intratumoral heterogeneity. J Cancer 2016; 7(13): 1907–1914.
15. Diaz LA Jr, Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. J Clin Oncol 2014; 32(6): 579–586.
16. Cancer Genome Atlas Research Network. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. Cell 2017; 169: 1327–1341 e1323.
17. Cleary SP, Jeck WR, Zhao X et al. Identification of driver genes in hepatocellular carcinoma by exome sequencing. Hepatology 2013; 58(5): 1693–1702.
18. Fujimoto A, Furuta M, Totoki Y et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. Nat Genet 2016; 48(5): 500–509.
19. Fujimoto A, Totoki Y, Abe T et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. Nat Genet 2012; 44(7): 760–764.
20. Guichard C, Amaddeo G, Imbeaud S et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. Nat Genet 2012; 44(6): 694–698.
21. Kan Z, Zheng H, Liu X et al. Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. Genome Res 2013; 23(9): 1422–1433.
22. Weinhold N, Jacobsen A, Schultz N et al. Genome-wide analysis of noncoding regulatory mutations in cancer. Nat Genet 2014; 46(11): 1160–1165.
23. Llovet JM, Bru C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. Semin Liver Dis 1999; 19(03): 329–338.
24. Edmondson HA, Steiner PE. Primary carcinoma of the liver: a study of 100 cases among 48, 900 necropsies. Cancer 1954; 7(3): 462–503.
25. Chakravarty D, Gao J, Phillips SM et al. OncoKB: a precision oncology knowledge base. JCO Precis Oncol 2017; 1–16.

9.3 Phosphorylated CXCR4 expression has a positive prognostic impact in colorectal cancer.

Cell Oncol. (2017) 40:609–619
DOI 10.1007/s13402-017-0348-2

ORIGINAL PAPER

Phosphorylated CXCR4 expression has a positive prognostic impact in colorectal cancer

B. Weixler[1] · F. Renetseder[1] · I. Facile[1] · N. Tosti[2] · E. Cremonesi[3] · A. Tampakis[1] · T. Delko[1] · S. Eppenberger-Castori[2] · A. Tzankov[2] · G. Iezzi[3] · C. Kettelhack[1] · S. D. Soysal[1] · U. von Holzen[4,5] · G. C. Spagnoli[3] · L. Terracciano[2] · L. Tornillo[2] · Raoul A. Droeser[1] · S. Däster[1]

Accepted: 4 September 2017 / Published online: 21 September 2017

Abstract
Background The CXCL12-CXCR4 chemokine axis plays an important role in cell trafficking as well as in tumor progression. In colorectal cancer (CRC), the chemokine receptor CXCR4 has been shown to be an unfavorable prognostic factor in some studies, however, the role of its activated (phosphorylated) form, pCXCR4, has not yet been evaluated. Here, we aimed to investigate the prognostic value of CXCR4 and pCXCR4 in a large cohort of CRC patients.
Patients and methods A tissue microarray (TMA) of 684 patient specimens of primary CRCs was analyzed by immunohistochemistry (IHC) for the expression of CXCR4 and pCXCR4 by tumor cells and tumor-infiltrating immune cells (TICs).
Results The combined high expression of CXCR4 and pCXCR4 showed a favorable 5-year overall survival rate (68%; 95%CI = 59–76%) compared to tumors showing a high expression of CXCR4 only (48%; 95%CI = 41–54%). High expression of pCXCR4 was significantly associated with a favorable prognosis in a test and validation group ($p = 0.015$ and $p = 0.0001$). Moreover, we found that CRCs with a high density of pCXCR4+ tumor-infiltrating immune cells (TICs) also showed a favorable prognosis in a test and validation group ($p = 0.054$ and $p = 0.004$). Univariate Cox regression analysis for TICs revealed that a high density of pCXCR4+ TICs was a favorable prognostic marker for overall survival (HR = 0.97,95%CI = 0.96–1.00; $p = 0.01$). In multivariate Cox regression survival analyses a high expression of pCXCR4 in tumor cells lost its association with a better overall survival (HR = 0.99; 95%CI = 0.99–1.00, $p = 0.098$).
Conclusion Our results show that high densities of CXCR4 and pCXCR4 positive TICs are favorable prognostic factors in CRC.

Keywords CXCR4 · pCXCR4 · Colorectal cancer · Prognostic significance · CRC

1 Introduction

Colorectal cancer (CRC) is the second most common cause of cancer mortality in developed countries [1]. While surgical tumor resection remains the cornerstone therapy for localized CRC, adjuvant therapy is recommended for patients with locally advanced stages and palliative systemic therapy for patients with metastatic disease [2]. The selection for adjuvant chemotherapy is largely based on clinical criteria such as tumor node metastasis (TNM) classification and other histopathological factors. However, these clinico-pathological features alone may inadequately predict cancer aggressiveness [3, 4]. Therefore, several prognostic and predictive biomarkers have been studied to improve prognostic information and patient selection for adjuvant treatment [5, 6]. Numerous studies have conclusively shown that microenvironmental factors are crucial for CRC development and progression [7, 8] and also highlight the role of chemokines in tumor invasion and cancer metastasis [9]. As metastatic disease dramatically decreases

B. Weixler and F. Renetseder have equally contributed to this work

✉ Raoul A. Droeser
Raoul.Droeser@usb.ch

[1] Department of Surgery, University Hospital Basel, Basel, Switzerland
[2] Institute of Pathology, University of Basel, Basel, Switzerland
[3] Department of Biomedicine, University of Basel, Basel, Switzerland
[4] Goshen Center for Cancer Care, Indiana University School of Medicine South Bend, Goshen, IN, USA
[5] Harper Cancer Research Institute, South Bend, IN, USA

survival [10], it is pivotal to continue evaluating the prognostic value of chemokines in CRC.

The chemokine CXCL12 and one of its receptors, CXCR4, have been shown to play key roles in the tumor-stromal communication affecting cancer growth, angiogenesis and metastasis formation [11]. CXCR4 expression has been linked to cancer progression and metastasis in hematopoietic as well as in various non-hematopoietic malignancies [9, 12, 13]. It has also been reported as a prognostic marker in cancers of different origin [14–16]. Its interaction with CXCL12 is thought to play an important role in tumor proliferation, invasion, lymph node homing and metastatic progression. After binding by the corresponding ligand CXCL12, CXCR4 is phosphorylated [17]. Since the phosphorylated form of CXCR4 (pCXCR4) plays the biologically active role, the analysis of CXCR4 expression alone appears to be insufficient to support its functional role in cancer metastasis.

Literature regarding the prognostic role of CXCR4 is relatively vague. Some studies have shown that high CXCR4 expression in CRC patients correlates with an advanced tumor stage [18], an increased risk for recurrence and distant metastasis [19–21] and a poor overall survival [22]. A recent meta-analysis concluded that there is a significant association between CXCR4 expression and poor survival [23]. However, most studies have included small numbers of patients and have presented important methodological differences. On the other hand, there are also studies that failed to observe significant correlations between CXCR4 expression and metastasis [24] and/or overall survival [17, 25].

pCXCR4 expression has been reported to have a prognostic value superior to that of CXCR4 expression in breast cancer [26]. Interestingly, there is a lack of data regarding the prognostic significance of pCXCR4 in CRC. In addition, most studies have only evaluated CXCR4 expression on tumor cells, while its expression on tumor-infiltrating immune cells (TICs) has not been thoroughly investigated.

The purpose of our study was to comparatively investigate the prognostic value of CXCR4 and pCXCR4 expression in a large cohort of CRC patients. In addition to CXCR4/pCXCR4 expression by the tumor cells, we also assessed the relevance of CXCR4/pCXCR4 positive TICs.

2 Materials and methods

2.1 Tissue microarray construction

684 unselected, non-consecutive, clinically annotated, primary CRC specimens were included in a tissue microarray (TMA) following approval by the local ethics committee. Formalin-fixed, paraffin-embedded (FFPE) tissue blocks were prepared according to standard procedures. Tissue cylinders with a diameter of 0.6 mm were punched from morphologically representative areas of each donor block and brought into one recipient paraffin block (30 × 25mm), using a semi-automated tissue arrayer. Each punch was made from the center of the tumor to enable each TMA spot to include at least 50% tumor cells.

2.2 Clinico-pathological features

Clinico-pathological data were collected retrospectively in a non-stratified and non-matched manner. Annotation included patient age and gender, tumor diameter, location, pT/pN stage, grade, histologic subtype, vascular invasion, border configuration, presence of peritumoral lymphocytic inflammation at the invasive tumor front and disease-specific survival. Tumor border configuration and peritumoral lymphocytic inflammation were evaluated using the original H&E slides of the resection specimens corresponding to each TMA punch.

2.3 Immunohistochemistry

Staining protocols for the primary antibodies directed against CXCR4 (Abcam, ab2074; 1:50) and pCXCR4 (Abcam, ab74012; 1:200) were performed as recommended by the manufacturers, including positive control tissue samples exactly as previously described [27]. Immunohistochemistry was performed using the automated staining system Benchmark XT (Roche/Ventana Medical Systems, Tuscon, AZ).

2.4 Evaluation of immunohistochemistry

Immunohistochemical readings were performed by two trained research fellows [F.R. and I.F] and data were independently validated by an additional investigator [L.T.]. Histoscores for tumor cells were obtained by multiplying percentages positive cells by stain intensities (0 = negative, 1 = weak, 2 = moderate, 3 = strong). TICs were counted for each punch (approximately one high power [20×] field).

2.5 Statistical analysis

All statistical analyses were performed using STATA software version 13 (StataCorp, College Station, TX, USA). Associations with survival were explored using the Cox proportional hazards regression model. Cut-off values used to classify CRC with low or high immune cell infiltrations were obtained using median values. Therefore, threshold values for CXCR4 and pCXCR4 positivity in TICs were 23 and 2 cells/TMA-punch and 100 and 0 for CXCR4 and pCXCR4 for histoscores, related to tumor expression, respectively. Chi-square, Fisher's exact, and Kruskal-Wallis tests were used to determine associations between CXCR4 and pCXCR4 positivity and clinico-pathological features.

For survival analysis, the study population was randomly assigned to test and validation groups. Univariate survival analysis was performed by Kaplan-Meier and log rank tests. Further analysis included four combinations of CXCR4 and pCXCR4 positivity: CXCR4−/pCXCR4-, CXCR4+/pCXCR4-, CXCR4−/pCXCR4+ and CXCR4+/pCXCR4 + .

The assumption of proportional hazards was verified for all markers by analyzing correlation of Schoenfeld residuals and ranks of individual failure times. Any missing clinicopathological information was assumed to be missing at random. Subsequently, CXCR4 and pCXCR4 data were entered into multivariate Cox regression analysis and hazard ratios (HR), and 95% confidence intervals (CI) were used to determine prognostic effects on survival time. *P*-values < 0.05 were considered statistically significant.

3 Results

3.1 Patient and tumor characteristics

Tissue samples from a total of 684 CRC patients were analyzed. The median age was 70 years (range: 30–95) and 53.2% of the patients were female. In 69.3% of the patients CRC was located in the left hemicolon and in the remaining 30.4% in the right hemicolon. The TMA included 600 mismatch repair (MMR)-proficient CRC specimens and 84 MMR-deficient CRC specimens (12.3%), as identified by MLH1, MSH2 and MSH6 expression analysis [28]. The median overall survival was 55 months (range 0–151) and the 5-year overall survival rate was 53.7 (95% CI = 49.8–57.4). Losses due to missing information or miscarried TMA punches usually represented about 15% of the data (Table 1).

Table 1 Clinico-pathological characteristics of the overall CRC patient cohort (n = 684*)

Characteristics	N or mean	(% or range)
Age, years (median, mean)	70, 69.2	(30–95)
Tumor size in mm (median, mean)	50, 50.7	(4–170)
Sex		
Female (%)	364	53.2
Male (%)	320	46.8
Anatomic site of the tumor		
Left-sided (%)	474	69.3
Right-sided (%)	208	30.4
T stage		
T1 (%)	35	5.1
T2 (%)	98	14.3
T3 (%)	444	64.9
T4 (%)	94	13.7
N stage		
N0 (%)	363	53.1
N1 (%)	168	24.6
N2 (%)	137	20.0
Tumor grade		
G1 (%)	21	3.1
G2 (%)	606	88.6
G3 (%)	43	6.3
UICC		
Stage IA (%) T1 N0	28	4.1
Stage IB (%) T2 N0	71	10.4
Stage IIA (%) T3 N0	230	33.6
Stage IIB-C (%) T4 N0	28	4.1
Stage III (%) N+	300	43.9
Tumor border configuration		
Infiltrative (%)	464	67.8
Pushing (%)	205	30.0
Vascular invasion		
No (%)	484	70.8
Yes (%)	187	27.3
Microsatellite Stability		
Proficient (%)	600	87.7
Deficient (%)	84	12.3
CXCR4 histoscore (mean, SD)	154.6	(±98.7)
pCXCR4 histoscore (mean, SD)	31.7	(±62.4)
CXCR4 TICI (mean, SD)	76.3	(±153.2)
pCXCR4 TICI (mean, SD)	5.1	(±10.2)
Median overall survival time (months)	55	0–151
5-year overall survival % (95%CI)	53.7	49.8–57.4

*Percentages may not add to 100% due to missing values of defined variables

3.2 CXCR4 and pCXCR4 expression in CRC surgical specimens

The expression of CXCR4 and pCXCR4 in tumor cells and in TICs was highly variable. Representative examples of negative and positive CXCR4/pCXCR4 tumor cells and TICs are presented in Fig. 1. Expectably, only a fraction of CXCR4+ cells showed evidence of receptor phosphorylation. The mean values for CXCR4 and pCXCR4 histoscores were 154.6 (± 98.7) and 31.7 (± 62.4), respectively. For TIC density, the corresponding numbers were 76.3 (±153.2) and 5.1 (±10.2), respectively.

High correlation coefficients were observed between CXCR4+ and pCXCR4+ immune cell infiltration (r_s = 0.363, p < 0.001) and CXCR4 and pCXCR4 histoscores (r_s = 0.264, p < 0.001), thus confirming the integrity of our measurements. More importantly, pCXCR4 histoscore and pCXCR4+ immune cell infiltration were also significantly correlated (r_s = 0.244, p < 0.001), suggesting that the CXCL12 ligand may be active in the tumor microenvironment on both tumor and infiltrating immune cells. Instead, we found that CXCR4+ histoscore and CXCR4+ immune cell infiltration were poorly correlated (r_s = 0.121).

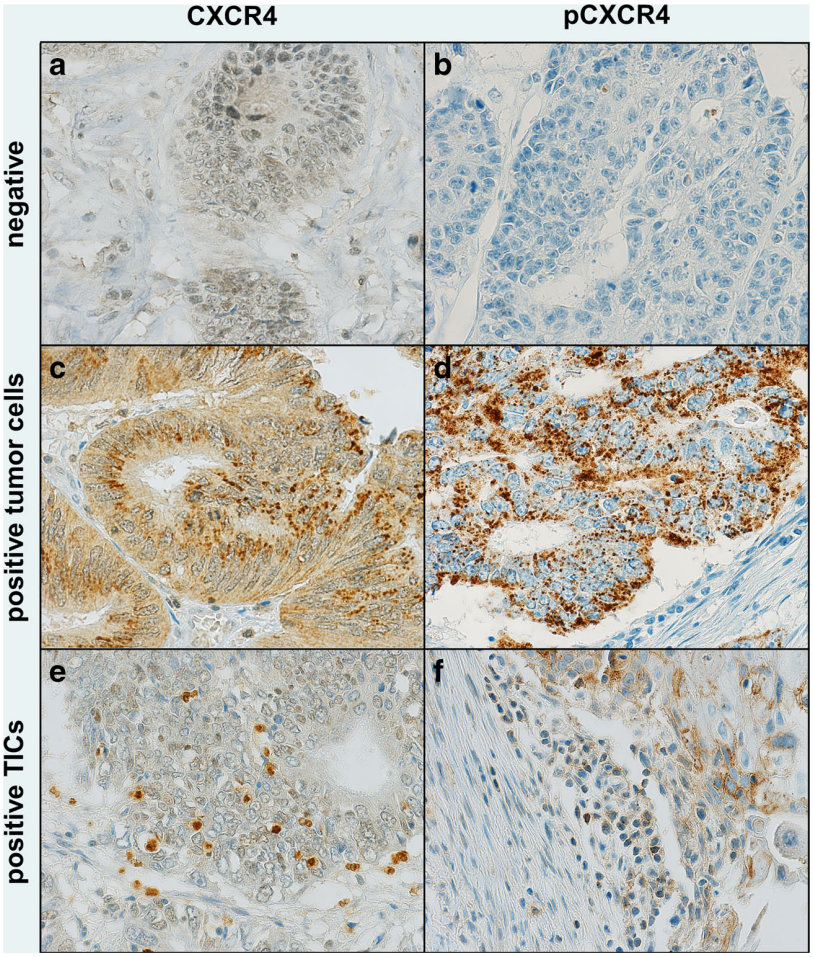


Fig. 1 Positive (high) CXCR4/pCXCR4 expression in CRC cells and CRC infiltration by CXCR4/pCXCR4-positive TICs compared to negative samples. CRC samples were stained with CXCR4 and pCXCR4-specific reagents. Tumor punches are representative of absence of infiltration (panel **a** and **b**), positive/high expression on tumor cells (panel **c** and **d**), and positive/high infiltration of CXCR4/pCXCR4-positive TICs (panels **e** and **f**), respectively. Magnification: 60×

3.3 Association of clinico-pathological features according to CXCR4 and pCXCR4 expression

Clinico-pathological features under investigation and their relation to the four subgroups identified using median values of CXCR4+ and pCXCR4+ expression (CXCR4low/pCXCR4low, CXCR4high/pCXCR4low, CXCR4low/pCXCR4high and CXCR4high/pCXCR4high) are listed in Tables 2 and 3 for histoscores and immune cell density infiltration, respectively. CRC tissues with a high expression of pCXCR4, alone or in combination with CXCR4 in the tumor cells, were significantly more frequently located in the left-sided colon ($p = 0.006$) and were characterized by a significantly lower T- ($p = 0.001$) and N-stage ($p < 0.001$), absence of vascular invasion ($p = 0.004$) and absence of an infiltrative tumor border ($p = 0.001$) (Table 2). CRCs with a high density of CXCR4/pCXCR4+ or pCXCR4+ TICs were also characterized by a significantly lower N-stage ($p < 0.001$) (Table 3).

In order to directly evaluate the relevance of receptor phosphorylation, we next differentially analyzed CRC clinico-pathological features in tissues with a high CXCR4 but a different pCXCR4 expression. Most interestingly, we found that CRCs with a high CXCR4+ histoscore were characterized by a significantly more frequent pT1-2 and pN0 stage in pCXCR4+ compared to pCXCR4- cases ($p = 0.011$ and $p = 0.002$, respectively) and by a significantly more frequent detection of a "pushing" tumor border ($p = 0.028$).

Regarding TICs in the presence of a high CXCR4+ infiltration, we found that CRCs displaying pCXCR4+ "high" infiltrates were significantly more frequently characterized by a pN0 stage compared to tumors with pCXCR4+ immune cell infiltration ($p = 0.0004$). Taken together, these data suggest that receptor phosphorylation, rather than mere receptor expression, is associated with important clinico-pathological characteristics.

3.4 Prognostic significance of CXCR4 and pCXCR4 expression by tumor cells and by tumor-infiltrating immune cells

Kaplan-Meier curves related to histoscore data revealed that pCXCR4 expression was significantly associated with a favorable prognosis in both test and validation groups ($p = 0.0001$ and $p = 0.015$; Fig. 2a and b). In sharp contrast,

Table 2 Association of CXCR4+ and pCXCR4+ low and high expression by tumor cells (histoscore) and clinico-pathological features in CRC (cut-offs were 100 and 0 for CXCR4 and pCXCR4, respectively)

		$CXCR4^{high}$/ $pCXCR4^{high}$		$CXCR4^{low}$/ $pCXCR4^{high}$		$CXCR4^{high}$/ $pCXCR4^{low}$		$CXCR4^{low}$/ $pCXCR4^{low}$		*p*-value
		N = 120	(100%)	*N* = 75	(100%)	*N* = 147	(100%)	*N* = 214	(100%)	
Age	years, mean ± SD	68.3	11.6	69.8	11.0	70.5	10.3	68.2	11.5	0.322
Tumor diameter	mm, mean ± SD	49.2	21.4	45.7	17.3	54.2	20.6	50.8	19.5	**0.009***
Gender	Female	67	55.8	40	53.3	83	56.5	104	43.2	0.433
	Male	53	44.2	35	46.7	64	43.5	110	45.6	
Tumor location	Left-sided	87	72.5	65	86.7	103	70.1	140	58.1	**0.006****
	Right-sided	33	27.5	10	13.3	44	29.9	72	29.9	
Histologic subtype	Mucinous	5	4.2	2	2.7	10	6.8	7	2.9	0.842
	Non-mucinous	115	95.8	73	97.3	137	93.2	207	85.9	
pT stage	pT1	12	10	8	10.7	8	5.4	3	1.2	**0.001*****
	pT2	24	20	13	17.3	16	10.9	27	11.2	
	pT3	72	60	45	60	100	68	147	61	
	pT4	11	9.2	6	8	21	14.3	32	13.3	
pN stage	pN0	84	70	45	60	74	50.3	96	39.8	**<0.001******
	pN1	17	14.2	18	24	42	28.6	59	24.5	
	pN2	17	14.2	10	13.3	25	17	55	22.8	
Tumor grade	G1	6	5	4	5.3	4	2.7	4	1.7	0.268
	G2	106	88.3	65	86.7	133	90.5	185	76.8	
	G3	7	5.8	3	4	7	4.8	20	8.3	
Vascular invasion	Absent	97	80.8	56	74.7	107	72.8	134	55.6	**0.004**†
	Present	22	18.3	16	21.3	38	25.9	75	31.1	
Tumor border	Pushing	51	42.5	25	33.3	43	29.3	47	19.5	**0.001**‡
	Infiltrating	68	56.7	45	60	102	69.4	162	67.2	
PTL inflammation	Absent	92	76.7	54	72	112	76.2	162	67.2	0.974
	Present	27	22.5	18	24	33	22.4	47	19.5	
Microsatellite stability	Deficient	7	5.8	8	10.7	16	10.9	31	12.7	0.111
	Proficient	113	94.2	67	89.3	131	89.1	183	75.9	
5-year overall survival rate	95%CI	0.68	0.59–0.76	0.63	0.50–0.73	0.48	0.39–0.56	0.48	0.41–0.54	**<0.001**§

Percentages may not add to 100% due to missing values of defined variables. Variables are indicated as absolute numbers, %, median or range. Age and tumor size were evaluated using the Kruskal-Wallis test. Gender, anatomical site, T stage, N stage, grade, vascular invasion, and tumor border configuration were analyzed using the χ^2 test. Survival analysis was evaluated using the Kaplan-Meier method. *P*-values < 0.05 are highlighted in bold. Significant combinations of CXCR4/pCXCR4 in pairwise analyses:

*− + vs + − ($p = 0.003$), ++ vs + − ($p = 0.044$);

**− + vs + − ($p = 0.008$), ++ vs − + ($p = 0.022$), – vs − + ($p = 0.001$);

***– vs − + ($p = 0.002$), – vs ++ ($p < 0.001$);

****++ vs + − ($p = 0.005$), – vs − + ($p = 0.036$), – vs ++ ($p < 0.001$);

† – vs − + ($p = 0.041$), – vs ++ ($p = 0.001$);

‡ ++ vs + − ($p = 0.029$), – vs − + (p = 0.04), – vs ++ ($p < 0.001$);

§ − + vs + − ($p = 0.006$), ++ vs + − ($p < 0.001$), – vs − + ($p = 0.003$), – vs ++ ($p < 0.001$)

we found that expression of CXCR4 in the absence of pCXCR4 had no impact on overall survival, even considering the whole cohort ($p = 0.132$; data not shown). Indeed, CRCs with a high expression of CXCR4 and pCXCR4 on tumor cells showed a favorable 5-year overall survival rate (68%, 95%CI 59–76%) ($p < 0.001$; Table 2), whereas no difference was observed between tumors with a high CXCR4 expression only (48%, 95%CI = 39–56%) or low CXCR4/pCXCR4 expression (48%, 95%CI = 41–54%).

CRC with a high density of CXCR4/pCXCR4+ infiltrating TICs also showed a more favorable 5-year overall survival rate (63%, 95%CI = 54%-70%) compared to tumors with a low CXCR4/pCXCR4+ TIC density ($p = 0.004$) (Table 3). Kaplan-Meier curves for TICs revealed a trend to a more favorable prognosis for CRC with a high density of pCXCR4+ TICs compared to tumors with a low pCXCR4+ TIC density in a test group ($p = 0.054$; Fig. 2d) and a significant association with a more favorable prognosis in a

Table 3 Association of CXCR4+ and pCXCR4+ low and high immune cell density infiltration and clinico-pathological features in CRC (cut-offs were 23 and 2 cells for CXCR4 and pCXCR4, respectively)

		$CXCR4^{high}$/ $pCXCR4^{high}$		$CXCR4^{low}$/ $pCXCR4^{high}$		$CXCR4^{high}$/ $pCXCR4^{low}$		$CXCR4^{low}$/ $pCXCR4^{low}$		*p*-value
		N = 155	(100%)	*N* = 81	(100%)	*N* = 122	(100%)	*N* = 195	(100%)	
Age	years, mean ± SD	69.7	11.4	69.8	11.8	67.9	11.7	68.9	10.4	0.481
Tumor diameter	mm, mean ± SD	49.7	19.4	51.3	25.1	50.6	18.2	51.4	19.6	0.893
Gender	Female	80	51.6	48	59.3	61	50	102	52.3	0.609
	Male	75	48.4	33	40.7	61	50	93	47.7	
Tumor location	Left-sided	115	74.2	57	70.4	85	69.7	135	69.2	0.713
	Right-sided	39	25.2	24	29.6	36	29.5	60	30.8	
Histologic subtype	Mucinous	8	5.2	2	2.5	7	5.7	7	3.6	0.477
	Non-mucinous	147	94.8	79	97.5	115	94.3	188	96.4	
pT stage	pT1	18	11.6	4	4.9	5	4.1	4	2.1	0.068
	pT2	21	13.5	14	17.3	20	16.4	24	12.3	
	pT3	94	60.6	52	64.2	77	63.1	139	71.3	
	pT4	18	11.6	9	11.1	19	15.6	24	12.3	
pN stage	pN0	103	66.5	45	55.6	58	47.5	90	46.2	**<0.001***
	pN1	24	15.5	19	23.5	33	27	61	31.3	
	pN2	21	13.5	15	18.5	30	24.6	40	20.5	
Tumor grade	G1	8	5.2	3	3.7	1	0.8	6	3.1	0.464
	G2	134	86.5	69	85.2	111	91	171	87.7	
	G3	8	.5.2	7	8.6	9	7.3	14	7.2	
Vascular invasion	Absent	120	77.4	57	70.4	83	68	131	67.2	0.104
	Present	31	20	22	27.2	38	31.1	60	30.8	
Tumor border	Pushing	52	33.5	28	34.6	32	26.2	53	27.2	0.292
	Infiltrating	99	63.9	50	61.7	89	72.9	137	70.3	
PTL inflammation	Absent	112	72.3	56	69.1	94	77.1	156	80	0.184
	Present	39	25.2	23	28.4	27	22.1	35	17.9	
Microsatellite stability	Deficient	10	6.5	11	13.6	16	13.1	25	12.8	0.142
	Proficient	145	93.5	70	86.4	106	86.9 g	170	87.2	
5-year overall survival rate	95%CI	0.63	0.54-0.70	0.63	0.51-0.72	0.49	0.40-0.58	0.46	0.38-0.52	**0.004****

Percentages may not add to 100% due to missing values of defined variables. Variables are indicated as absolute numbers, %, median or range. Age and tumor size were evaluated using the Kruskal-Wallis test. Gender, anatomical site, T stage, N stage, grade, vascular invasion, and tumor border configuration were analyzed using the χ^2 test. Survival analysis was evaluated using the Kaplan-Meier method. *P*-values < 0.05 are highlighted in bold. Significant combinations of CXCR4/pCXCR4 in pairwise analyses:

*++ vs – ($p < 0.001$), ++ vs + − ($p = 0.002$);

**++ vs – ($p < 0.001$)

validation group ($p = 0.004$; Fig. 2e). Kaplan-Meier curves for tumors with CXCR4/pCXCR4 high TIC infiltrates compared to tumors with CXCR4 high pCXCR4 low TIC densities showed a trend to a better prognosis for the CXCR4 high/ pCXCR4 high group ($p = 0.06$, Fig. 2f).

3.5 Univariate and multivariate analysis of CXCR4 and pCXCR4 expression by tumor cells and tumor-infiltrating immune cells

Univariate Cox regression analysis of the histoscores revealed that CXCR4 and pCXCR4 positive staining is significantly associated with an increased overall survival (HR 0.99; 95%CI = 0.99–1.0; $p = 0.035$ and $p < 0.001$, respectively). Age, male gender, T-stage, tumor grading, N-stage, invasive margin and vascular invasion were all significantly associated with a poor prognosis in univariate analyses (Table 4). Univariate Cox regression analysis for TICs showed that a high density of pCXCR4+ TICs also serves as a favorable prognostic marker for overall survival (HR = 0.98; 95%CI = 0.97–1.00; $p = 0.02$). Instead, we found that the density of CXCR4+ TICs had no impact on overall survival (Table 5).

Through multivariate hazard Cox regression survival analysis, however, we found that high expression of pCXCR4+ on tumor cells (histoscore) failed to retain its role as an independent prognostic factor for overall survival (HR = 0.99; 95%CI = 0.99–1.0; $p = 0.098$; Table 4). Similarly, multivariate

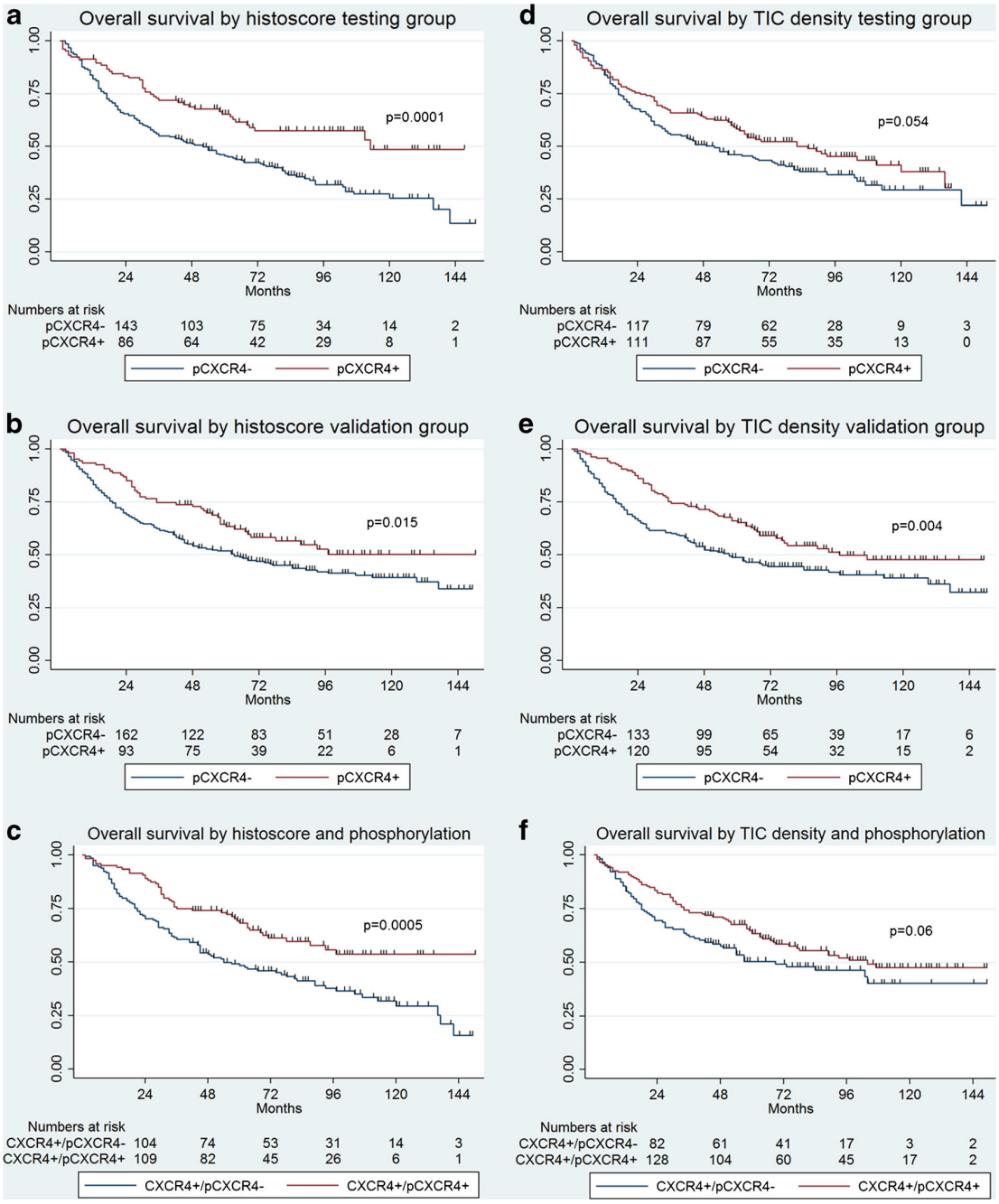


Fig. 2 Effects of pCXCR4+ tumor cells and pCXCR4+ TIC infiltration on overall survival in patients with CRC. Kaplan-Meier curves were created according to the expression of pCXCR4 on tumor cells (curves **a**-**c**) and pCXCR4+ TICs (curves **d**-**f**) in patients with CRC (n = 684). Kaplan-Meier overall survival curves for the histoscore show that high expression of pCXCR4 was significantly associated with a favorable prognosis in a test (**a**, p = 0.0001) and validation (**b**, p = 0.015) group. Overall survival was also evaluated for CRC showing evidence of a CXCR4/pCXCR4 high histoscore, as compared to tumors with CXCR4 high pCXCR4 low values on the whole cohort (**c**, p = 0.005). Kaplan-Meier curves for infiltration by TICs showed at least a trend towards a favorable prognosis for CRCs with a high density of pCXCR4+ TICs in a test (**c**, p = 0.054) and validation (**d**, p = 0.004) group. Kaplan-Meier curves were also analyzed for CRC showing evidence of CXCR4/pCXCR4 high TIC infiltrates, as compared to tumors with CXCR4 high pCXCR4 low TIC densities on the whole cohort (**f**, p = 0.06)

Hazard Cox regression survival analysis of TICs showed the absence of an independent prognostic significance for CXCR4+ (HR = 1.0; 95%CI = 0.99–1.0; p = 0.920) and pCXCR4+ (HR = 0.99; 95%CI = 0.97–1.01; p = 0.200) immune cell infiltration on overall survival (Table 5). On the other hand, we found that an increased age (HR = 1.04; 95%CI = 1.03–1.05; p < 0.001), male gender (HR = 1.53; 95%CI = 1.2–1.9; p < 0.001), a higher T-stage (HR = 2.02;

Table 4 Uni- and multivariate Hazard Cox regression survival analysis by histoscore

	Univariate			Multivariate		
	HR	95% CI	*p*-values	HR	95% CI	*p*-values
CXCR4 histoscore	0.99	0.99–1.0	**0.035**	0.99	0.99–1.0	0.322
pCXCR4 histoscore	0.99	0.99–1.0	**<0.001**	0.99	0.99–1.0	0.098
Age	1.02	1.01–1.03	**<0.001**	1.04	1.03–1.05	**<0.001**
Gender (men vs women)	1.32	1.07–1.61	**0.008**	1.53	1.2–1.9	**<0.001**
pT stage (T3-4 vs T1-2)	2.91	2.08–4.08	**<0.001**	2.02	1.33–3.08	**0.001**
Tumor grade (high vs low)	5.56	1.78–17.32	**0.003**	2.45	0.77–7.8	0.129
pN stage (pos. vs neg.)	3.09	2.50–3.83	**<0.001**	2.38	1.85–3.07	**<0.001**
Invasive margin	1.94	1.52–2.48	**<0.001**	1.36	1.01–1.83	**0.043**
Vascular invasion	2.44	1.97–3.01	**<0.001**	1.64	1.27–2.12	**<0.001**

Uni- and multivariate Cox-regression analyses showing Hazard Ratios and *P*-values (*P*-values < 0.05 are highlighted in bold). The multivariate model included 524 patients, due to missing values related to CXCR4+ and pCXCR4+ tumor expression, age, sex, tumor grade, vascular invasion, tumor border configuration, pT and pN stage

95%CI = 1.33–3.08; $p = 0.001$) and N-stage (HR = 2.38; 95%CI = 1.85–3.07; $p < 0.001$) were independently associated with a poor prognosis (Tables 4 and 5).

4 Discussion

In the past, several studies have been carried out to evaluate the prognostic significance of CXCR4 expression in CRC. However, methodological differences, inappropriate immunohistochemical protocols and limited sample sizes have hampered the drawing of definite conclusions. Most importantly, none of these studies has analyzed the expression of the phosphorylated, activated form of this receptor, pCXCR4. To the best of our knowledge, this is the first study analyzing the prognostic significance of pCXCR4 in a large cohort of CRC patients. Our data indicate that CXCR4 expression in the absence of phosphorylation on tumor cells has no prognostic significance, whereas expression of its phosphorylated form pCXCR4 is associated with a favorable clinical outcome.

These data are conflicting with those of most previous studies, showing a negative prognostic impact of CXCR4 expression in CRC cells on survival. One study reported a negative prognostic impact of CXCR4 expression in stage I, II and IV CRCs [29]. Another study reported that high CXCR4 expression is associated with higher TNM stages, rectal cancer, metastases and a decreased survival [19]. However, very small sample collections ($n = 92$ and $n = 97$, respectively) question the significance of these analyses. High nuclear expression of CXCR4 has also been proposed to be associated with a poor survival in stage III CRC [30] and according to Spetjens et al. [31] nuclear, but not cytoplasmic, location of CXCR4 staining represents an independent negative prognostic factor. However, we failed to observe nuclear CXCR4 staining in our study cohort, and it has been reported that

Table 5 Uni- and multivariate Hazard Cox regression survival analysis of CXCR4 and pCXCR4 positive immune cell infiltration

	Univariate			Multivariate		
	HR	95% CI	*p*-values	HR	95% CI	*p*-values
CXCR4+ immune cells	0.99	0.99–1.00	0.424	1.00	0.99–1.00	0.920
pCXCR4+ immune cells	0.98	0.97–1.00	**0.019**	0.99	0.97–1.01	0.200
Age	1.02	1.01–1.03	**<0.001**	1.04	1.03–1.05	**<0.001**
Gender (men vs women)	1.32	1.07–1.61	**0.008**	1.58	1.24–2.00	**<0.001**
pT stage (T3-4 vs T1-2)	2.91	2.08–4.08	**<0.001**	2.11	1.38–3.24	**0.001**
Tumor grade (high vs low)	5.56	1.78–17.32	**0.003**	2.39	0.75–7.62	0.142
pN stage (pos. vs neg.)	3.09	2.50–3.83	**<0.001**	2.44	1.89–3.15	**<0.001**
Invasive margin	1.94	1.52–2.48	**<0.001**	1.38	1.03–1.87	**0.034**
Vascular invasion	2.43	1.97–3.01	**<0.001**	1.67	1.30–2.16	**<0.001**

Uni- and multivariate Cox-regression analyses showing Hazard Ratios and *P*-values (*P*-values < 0.05 are highlighted in bold). The multivariate model included 521 patients, due to missing values related to CXCR4+ and pCXCR4+ density, age, sex, tumor grade, vascular invasion, tumor border configuration, pT and pN stage

nuclear staining of CXCR4 seems to represent an artifact [32] related either to excessive antigen retrieval by pressure cooking or the application of poorly working primary antibodies, since the outer membrane receptor CXCR4 has no linked nuclear functions [33].

CXCL12, the only known ligand for CXCR4, is produced by a limited number of cell types including endothelial and bone marrow cells, mucosal epithelial cells, tumor cells and T-lymphocytes [34, 35]. CXCL12 expression is increased in tissues characterized by neo-angiogenesis and inflammation, supporting chemotactic gradients attracting CXCR4+ immune cells, mainly CD4+ and dendritic cells. As yet, the effects of CXCR4+ immune cells on tumor progression have not been studied in detail [13, 35]. Our study shows for the first time that activation of CXCR4, as suggested by the presence of its pCXCR4 form, in CRC tumors and infiltrating immune cells is significantly associated with a favorable prognosis. Current understanding of the CXCL12-CXCR4 axis postulates that CXCR4 expression by cancer cells guides them to migrate to ectopic sites with a high CXCL12 expression. However, our data indicate that the presence of pCXCR4 on tumor cells in CRC patients is associated with a lack of evidence of node metastases.

Consistent with our data, Stanisavljevic et al. [30] have shown that CXCL12 expression represents a positive prognostic factor for disease-free survival in CRC. Furthermore, Wendt et al. [36] showed that re-establishment of CXCL12 expression reduced metastasis in *in vivo* CRC experimental models. Remarkably, Roy et al. [37] also showed in an experimental model of pancreatic cancer that CXCL12 expression inhibited tumor growth and cancer cell metastasis formation through cell-cycle arrest, resulting in an increased overall survival.

From a mechanistic point of view, it is tempting to speculate that local CXCL12 production, resulting in CXCR4 phosphorylation, may facilitate the retention of CXCR4+ tumor cells within primary tumor tissues, thus preventing their migration towards potential metastatic sites. On the other hand, local CXCL12 production may favor the recruitment of immune cells associated with an improved prognosis, including CD8+ T-cells [3, 4, 38]. More intriguingly, a recent study has suggested that CXCL12 produced by neutrophils may guide activated CD8+ T-cells within mucosal tissues [39]. Previously, our group has observed that, indeed, neutrophil infiltration is also associated with a favorable prognosis in CRC [40]. However, CXCL12 production by these cells in CRC patients was not addressed. Further research is warranted to clarify the mechanistic aspects associated with the prognostic significance of pCXCR4 expression in CRC tissues.

To date, very few studies have been reported on the prognostic significance of pCXCR4. We could identify only two studies, one in B-acute lymphoblastic leukemia and another one in non-small cell lung cancer (NSCLC) [32, 41]. The study on NSCLC failed to detect any prognostic significance either for CXCR4 or pCXCR4 expression [32], whereas Konoplev et al. [41] reported a negative prognostic significance for pCXCR4 expression, but not CXCR4 expression, in B-acute lymphoblastic leukemia. However, consistent with our findings on total CXCR4 expression, Spano et al. [42] reported better disease outcomes for a higher CXCR4 expression in early-stage NSCLC. Importantly, the CXCL12-CXCR4 axis is considered a potential therapeutic target, not only in hematologic, but also in metastatic solid tumors [43] and clinical trials are now ongoing. Our data suggest an essential need of including the evaluation of both the presence of CXCR4 and its active form, pCXCR4, into the translational design of such studies. Another significant focus of the present study is the evaluation of CXCR4+ and pCXCR4+ TICs. Although there are many reports on the potential function of CXCR4 in different cancer types, there is a lack of data on its expression on such TICs. In this regard, our data may contribute to the definition of immune contexture features and their prognostic significance in CRC [4, 7, 38].

Our study suffers from a number of limitations. Firstly, the TMA technology may fail to represent tumor tissue heterogeneity. However, the punches included in this TMA were derived from tumor centers and included at least 50% cancer cells. Furthermore, the number of individual CRC specimens (> 600) compensates, at least in part, for the heterogeneity of the immune contexture in different tumor areas. Second, this is a retrospective cohort study. However, data emerging from large retrospective analyses may help in the development of prospective and mechanistic studies, currently planned by our group. Finally, the cohort investigated in this study includes CRC patients surgically treated between 1985 and 1998, prior to a widespread use of neoadjuvant treatment regimens for rectal cancer. Thus, while these results may not be fully representative of current clinical treatments, they are more likely to mirror CRC immunobiology. On the other hand, our results would urge the analysis of CXCL12 expression in the samples under investigation. However, detection of cytokines or, generally, soluble factors by immunohistochemistry is highly problematic and gene expression studies, although potentially suggestive, fail to provide evidence of specific protein production.

In summary, our data show for the first time that expression of activated pCXCR4 in tumor cells and a high density of CXCR4 and pCXCR4 positive immune cell infiltration in CRC represent favorable prognostic factors, thereby shedding a new light on the biological role of CXCR4 in CRC progression.

Acknowledgements GI was supported by SNF (CH) grant no. PP00P3-133699 and GCS was supported by SNF (CH) grant no. 310030_149745.

Compliance with ethical standards

Ethical statement All authors have agreed to the submission of this manuscript and have participated in the study to a sufficient extent to be named as authors.

Conflict of interest The authors have no conflicts of interest to disclose for the present study.

References

1. J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D.M. Parkin, D. Forman, F. Bray, Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer **136**, E359–E386 (2015)
2. H. Brenner, M. Kloor, C.P. Pox, Colorectal cancer. Lancet **383**, 1490–1502 (2014)
3. I. Zlobec, A. Lugli, Prognostic and predictive factors in colorectal cancer. Postgrad Med J **84**, 403–411 (2008)
4. F. Pagès, J. Galon, M.-C. Dieu-Nosjean, E. Tartour, C. Sautès-Fridman, W.-H. Fridman, Immune infiltration in human tumors: a prognostic factor that should not be ignored. Oncogene **29**, 1093–1102 (2010)
5. T. Tanaka, M. Tanaka, T. Tanaka, R. Ishigamori, Biomarkers for colorectal cancer. Int J Mol Sci **11**, 3209–3225 (2010)
6. G. Lech, R. Słotwiński, M. Słodkowski, I.W. Krasnodębski, Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances. World J Gastroenterol **22**, 1745–1755 (2016)
7. L.M. Coussens, Z. Werb, Inflammation and cancer. Nature **420**, 860–867 (2002)
8. A. Lasry, A. Zinger, Y. Ben-Neriah, Inflammatory networks underlying colorectal cancer. Nat Immunol **17**, 230–240 (2016)
9. J.A. Burger, CXCR4: a key receptor in the crosstalk between tumor cells and their microenvironment. Blood **107**, 1761–1767 (2006)
10. R. Siegel, J. Ma, Z. Zou, A. Jemal, Cancer statistics, 2014. CA Cancer J Clin **64**, 9–29 (2014)
11. F. Guo, Y. Wang, J. Liu, S.C. Mok, F. Xue, W. Zhang, CXCL12/CXCR4: a symbiotic bridge linking cancer cells and their stromal neighbors in oncogenic communication networks. Oncogene **35**, 816–826 (2015)
12. X. Sun, G. Cheng, M. Hao, J. Zheng, X. Zhou, J. Zhang, R.S. Taichman, K.J. Pienta, J. Wang, CXCL12 / CXCR4 / CXCR7 chemokine axis and cancer progression. Cancer Metastasis Rev **29**, 709–722 (2010)
13. L. Lombardi, F. Tavano, F. Morelli, T.P. Latiano, P. Di Sebastiano, E. Maiello, Chemokine receptor CXCR4: role in gastrointestinal cancer. Crit Rev Oncol Hematol **88**, 696–705 (2013)
14. C.C. Schimanski, P.R. Galle, M. Moehler, Chemokine receptor CXCR4-prognostic factor for gastrointestinal tumors. World J Gastroenterol **14**, 4721–4724 (2008)
15. B.E. Lippitz, Cytokine patterns in patients with cancer: A systematic review. Lancet Oncol **14**, e218–e228 (2013)
16. B.A. Teicher, S.P. Fricker, CXCL12 (SDF-1)/CXCR4 pathway in cancer. Clin Cancer Res **16**, 2927–2931 (2010)
17. M. Thelen, Dancing to the tune of chemokines. Nat Immunol **2**, 129–134 (2001)
18. S.-C. Wang, J.-K. Lin, H.-S. Wang, S.-H. Yang, A.F.-Y. Li, S.-C. Chang, Nuclear expression of CXCR4 is associated with advanced colorectal cancer. Int J Color Dis **25**, 1185–1191 (2010)
19. C.C. Schimanski, S. Schwald, N. Simiantonaki, C. Jayasinghe, U. Gönner, V. Wilsberg, T. Junginger, M.R. Berger, P.R. Galle, M. Moehler, Effect of chemokine receptors CXCR4 and CCR7 on the metastatic behavior of human colorectal cancer. Clin Cancer Res **11**, 1743–1750 (2005)
20. J. Kim, T. Mori, S.L. Chen, F.F. Amersi, S.R. Martinez, C. Kuo, R.R. Turner, X. Ye, A.J. Bilchik, D.L. Morton, D.S.B. Hoon, Chemokine receptor CXCR4 expression in patients with melanoma and colorectal cancer liver metastases and the association with disease outcome. Ann Surg **244**, 113–120 (2006)
21. N. Yoshitake, H. Fukui, H. Yamagishi, A. Sekikawa, S. Fujii, S. Tomita, K. Ichikawa, J. Imura, H. Hiraishi, T. Fujimori, Expression of SDF-1 alpha and nuclear CXCR4 predicts lymph node metastasis in colorectal cancer. Br J Cancer **98**, 1682–1689 (2008)
22. L.-N. Li, K.-T. Jiang, P. Tan, A.-H. Wang, Q.-Y. Kong, C.-Y. Wang, H.-R. Lu, J. Wang, Prognosis and Clinicopathology of CXCR4 in Colorectal Cancer Patients: a Meta-analysis. Asian Pac J Cancer Prev **16**, 4077–4080 (2015)
23. S. Lv, Y. Yang, S. Kwon, M. Han, F. Zhao, H. Kang, C. Dai, R. Wang, The association of CXCR4 expression with prognosis and clinicopathological indicators in colorectal carcinoma patients: a meta-analysis. Histopathology **64**, 701–712 (2014)
24. F. Xu, F. Wang, M. Di, Q. Huang, M. Wang, H. Hu, Y. Jin, J. Dong, M. Lai, Classification based on the combination of molecular and pathologic predictors is superior to molecular classification on prognosis in colorectal carcinoma. Clin Cancer Res **13**, 5082–5088 (2007)
25. S. Saigusa, Y. Toiyama, K. Tanaka, T. Yokoe, Y. Okugawa, A. Kawamoto, H. Yasuda, Y. Inoue, C. Miki, M. Kusunoki, Stromal CXCR4 and CXCL12 expression is associated with distant recurrence and poor prognosis in rectal cancer after chemoradiotherapy. Ann Surg Oncol **17**, 2051–2058 (2010)
26. S. Hassan, C. Ferrario, U. Saragovi, L. Quenneville, L. Gaboury, A. Baccarelli, O. Salvucci, M. Basik, The influence of tumor-host interactions in the stromal cell-derived factor-1/CXCR4 ligand/receptor axis in determining metastatic risk in breast cancer. Am J Pathol **175**, 66–73 (2009)
27. L. Brault, A. Rovó, S. Decker, C. Dierks, A. Tzankov, J. Schwaller, CXCR4-SERINE339 regulates cellular adhesion, retention and mobilization, and is a marker for poor prognosis in acute myeloid leukemia. Leukemia **28**, 566–576 (2014)
28. H. Hampel, J.A. Stephens, E. Pukkala, R. Sankila, L.A. Aaltonen, J.-P. Mecklin, A. de la Chapelle, Cancer risk in hereditary nonpolyposis colorectal cancer syndrome: later age of onset. Gastroenterology **129**, 415–421 (2005)
29. J. Kim, H. Takeuchi, S.T. Lam, R.R. Turner, H.-J. Wang, C. Kuo, L. Foshag, A.J. Bilchik, D.S.B. Hoon, Chemokine receptor CXCR4 expression in colorectal cancer patients increases the risk for recurrence and for poor survival. J Clin Oncol **23**, 2744–2753 (2005)
30. L. Stanisavljević, J. Aßmus, K.E. Storli, S.M. Leh, O. Dahl, M.P. Myklebust, CXCR4, CXCL12 and the relative CXCL12-CXCR4 expression as prognostic factors in colon cancer. Tumor Biol **37**, 7441–7452 (2015)
31. F.M. Speetjens, G.J. Liefers, C.J. Korbee, W.E. Mesker, C.J.H. Van De Velde, R.L. Van Vlierberghe, H. Morreau, R.A. Tollenaar, P.J.K. Kuppen, Nuclear localization of CXCR4 determines prognosis for colorectal cancer patients. Cancer Microenviron **2**, 1–7 (2009)
32. W. Sterlacci, S. Saker, B. Huber, M. Fiegl, A. Tzankov, Expression of the CXCR4 ligand SDF-1/CXCL12 is prognostically important for adenocarcinoma and large cell carcinoma of the lung. Virchows Arch **468**, 463–471 (2016)
33. T. Pozzobon, G. Goldoni, A. Viola, B. Molon, CXCR4 signaling in health and disease. Immunol Lett **177**, 6–15 (2016)
34. R.J. Phillips, M.D. Burdick, M. Lutz, J.A. Belperio, M.P. Keane, R.M. Strieter, The stromal derived factor-1/CXCL12-CXC chemokine receptor 4 biological axis in non-small cell lung cancer metastases. Am J Respir Crit Care Med **167**, 1676–1686 (2003)
35. O. Wald, U. Izhar, G. Amir, S. Avniel, Y. Bar-Shavit, H. Wald, I.D. Weiss, E. Galun, A. Peled, CD4+CXCR4highCD69+ T cells accumulate in lung adenocarcinoma. J Immunol **177**, 6983–6990 (2006)

36. M.K. Wendt, P.A. Johanesen, N. Kang-Decker, D.G. Binion, V. Shah, M.B. Dwinell, Silencing of epithelial CXCL12 expression by DNA hypermethylation promotes colonic carcinoma metastasis. Oncogene **25**, 4986–4997 (2006)
37. I. Roy, N.P. Zimmerman, A.C. Mackinnon, S. Tsai, D.B. Evans, M.B. Dwinell, CXCL12 chemokine expression suppresses human pancreatic cancer growth and metastasis. PLoS One **9**, e90400 (2014)
38. J. Galon, A. Costes, F. Sanchez-Cabo, A. Kirilovsky, B. Mlecnik, C. Lagorce-Pagès, M. Tosolini, M. Camus, A. Berger, P. Wind, F. Zinzindohoué, P. Bruneval, P.-H. Cugnenc, et al., Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science **313**, 1960–1964 (2006)
39. K. Lim, Y.-M. Hyun, K. Lambert-Emo, T. Capece, S. Bae, R. Miller, D.J. Topham, M. Kim, Neutrophil trails guide influenza-specific CD8+ T cells in the airways. Science **349**, aaa4352 (2015)
40. V. Governa, E. Trella, V. Mele, L. Tornillo, F. Amicarella, E. Cremonesi, M.G. Muraro, H. Xu, R. Droeser, S.R. Däster, M. Bolli, R. Rosso, D. Oertli, S. Eppenberger-Castori, L.M. Terracciano, G. Iezzi G, G.C. Spagnoli, The interplay between neutrophils and CD8(+) T cells improves survival in human colorectal cancer. Clin Cancer Res **23**, 3847–3858 (2017)
41. S. Konoplev, J.L. Jorgensen, D.A. Thomas, E. Lin, J. Burger, H.M. Kantarjian, M. Andreeff, L.J. Medeiros, M. Konopleva, Phosphorylated CXCR4 is associated with poor survival in adults with B-acute lymphoblastic leukemia. Cancer **117**, 4689–4695 (2011)
42. J.-P. Spano, F. Andre, L. Morat, L. Sabatier, B. Besse, C. Combadiere, P. Deterre, A. Martin, J. Azorin, D. Valeyre, D. Khayat, T. Le Chevalier, J.-C. Soria, Chemokine receptor CXCR4 and early-stage non-small cell lung cancer: pattern of expression and correlation with outcome. Ann Oncol **15**, 613–617 (2004)
43. A. Dubrovska, M. Cojoc, F. Peitzsch, F. Trautmann, G.D. Polishchuk, A. Telegeev, Emerging targets in cancer management: role of the CXCL12/CXCR4 axis. Onco Targets Ther **6**, 1347–1361 (2013)

9.4 HMGA1 expression in Human Hepatocellular Carcinoma correlated with poor prognosis and promotes tumor growth and migration in *in vitro* models.

NEOPLASIA
www.neoplasia.com

Volume 18 Number 12 December 2016 pp. 724–731 724

HMGA1 Expression in Human Hepatocellular Carcinoma Correlates with Poor Prognosis and Promotes Tumor Growth and Migration in *in vitro* Models[1]

Mariacarla Andreozzi*,[2], Cristina Quintavalle*,[2], David Benz*,[2], Luca Quagliata*, Matthias Matter*, Diego Calabrese[†], Nadia Tosti*, Christian Ruiz*, Francesca Trapani*, Luigi Tornillo*, Alfredo Fusco[‡,§], Markus H Heim[†], Charlotte KY Ng*, Pierlorenzo Pallante*,[‡], Luigi M Terracciano* and Salvatore Piscuoglio*

*Institute of Pathology, University Hospital Basel, Basel, Switzerland; [†]Department of Biomedicine, Hepatology Laboratory, University of Basel, Basel, Switzerland; [‡]Institute of Experimental Endocrinology and Oncology (IEOS), National Research Council (CNR), and Department of Molecular Medicine and Medical Biotechnology (DMMBM), University of Naples "Federico II, Naples, Italy; [§]National Cancer Institute-INCA, Rua Andrè Cavalcanti, 37-Centro, Rio de Janeiro, Brazil

Abstract

BACKGROUND: HMGA1 is a non-histone nuclear protein that regulates cellular proliferation, invasion and apoptosis and is overexpressed in many carcinomas. In this study we sought to explore the expression of HMGA1 in HCCs and cirrhotic tissues, and its effect in *in vitro* models. *METHODS:* We evaluated HMGA1 expression using gene expression microarrays (59 HCCs, of which 37 were matched with their corresponding cirrhotic tissue and 5 normal liver donors) and tissue microarray (192 HCCs, 108 cirrhotic tissues and 79 normal liver samples). HMGA1 expression was correlated with clinicopathologic features and patient outcome. Four liver cancer cell lines with stable induced or knockdown expression of HMGA1 were characterized using *in vitro* assays, including proliferation, migration and anchorage-independent growth. *RESULTS: HMGA1* expression increased monotonically from normal liver tissues to cirrhotic tissue to HCC ($P < .01$) and was associated with Edmondson grade ($P < .01$). Overall, 51% and 42% of HCCs and cirrhotic tissues expressed HMGA1, respectively. Patients with HMGA1-positive HCCs had earlier disease progression and worse overall survival. Forced expression of HMGA1 in liver cancer models resulted in increased cell growth and migration, and *vice versa*. Soft agar assay showed that forced expression of HMGA1 led to increased foci formation, suggesting an oncogenic role of HMGA1 in hepatocarcinogenesis. *CONCLUSIONS:* HMGA1 is frequently expressed in cirrhotic tissues and HCCs and its expression is associated with high Edmondson grade and worse prognosis in HCC. Our results suggest that HMGA1 may act as oncogenic driver of progression, implicating it in tumor growth and migration potential in liver carcinogenesis.

Neoplasia (2016) 18, 724–731

Introduction

HMGA1 is a non-histone nuclear protein involved in cell cycle-related chromosomal changes, genetic recombination, DNA replication and repair, apoptosis, and molecular chaperoning [1–4]. HMGA1 functions as an architectural transcriptional factor, as it regulates its target genes and microRNAs by direct DNA binding, forming transcriptional complexes and altering the conformation of transcription factors and chromatin structure [5–7]. HMGA1 is

Address all correspondence to: Salvatore Piscuoglio or Luigi Maria Terracciano, Institute of Pathology, University Hospital Basel, Schoenbeinstrasse 40, 4031 Basel, Switzerland.
E-mails: Luigi.Terracciano@usb.ch, Salvatore.Piscuoglio@usb.ch
[1] Funding: The study was supported by grants from Oncosuisse (KLS-3639-02-2015) and the Swiss Cancer Research foundation (KFS-3302-08-2013).
[2] Equally contributed to this work.
Received 8 September 2016; Revised 17 October 2016; Accepted 17 October 2016

1476-5586
http://dx.doi.org/10.1016/j.neo.2016.10.002

generally not expressed in adult tissues but is enriched in human embryonic and hematopoietic stem cells [8].

HMGA1 was first associated with the neoplastic phenotype in rat thyroid transformed cells [9] and has since been shown to lead to neoplastic transformation [3]. Of its many roles, HMGA1 negatively regulates *TP53* [10] and promotes an undifferentiated pluripotent stem-like cell state through the induction of *SOX2*, *LIN28* and *cMYC* [11]. HMGA1 also directly activates genes involved in tumor growth, migration, invasion, resistance to drug-induced cell death and epithelial-mesenchymal transition in cancer cells [12–15]. Indeed, HMGA1 overexpression has been reported in carcinomas of the colon, breast, pancreas, ovary, lung, esophagus and testis [16–25], and it correlated with advanced stage, the presence of distant metastases and poor survival in colorectal carcinomas [16,17]. Furthermore, HMGA1 expression levels have been found to increase progressively from no expression in normal breast tissue, to moderate expression in hyperplastic lesions to strong overexpression in ductal carcinomas [18], and to increase from weakly expressed in ovarian carcinomas with low invasive potential to be highly expressed in invasive carcinomas [21].

The *HMGA1* locus (6p21.3) is gained in around 40% of hepatocellular cancers (HCCs) [26], and an early study suggested that HMGA1 is expressed in 30% of primary HCC on the mRNA level and 13% on the protein level [27]. Furthermore, *HMGA1* mRNA expression was found to correlate with Edmondson grade and worse prognosis [27]. The functional significance of HMGA1 overexpression, however, has not been assessed. In this study, we evaluated *HMGA1* mRNA expression levels in a cohort of HCC needle biopsies matched with their corresponding cirrhotic tissues and normal liver donors by gene expression microarrays and quantitative real-time PCR (qRT-PCR). Using tissue microarray (TMA) technology, we further corroborated our results at the protein level in a large independent collection of 379 specimens including normal liver, cirrhotic and HCC tissues. Finally, we showed that HMGA1 overexpression promoted tumor growth and migration potential in liver carcinogenesis.

Materials and Methods

Ethics

The study has been approved by the Institutional Review Board of the Institute of Pathology, University Hospital, Basel and the Ethics Committee of Nordwest/Central Switzerland (EKNZ).

Re-analysis of Transcriptomic Profiling Data

HMGA1 expression was evaluated in 59 HCC needle biopsies, 37 of which were matched with their corresponding non-neoplastic liver parenchyma (cirrhotic tissues) and 5 normal liver donors using transcriptomic data our group previously published (GSE64041) [28]. CEL files were normalized using the Qlucore software (Qlucore AB, Lund, Sweden) [29]. *HMGA1* expression was extracted for each sample.

Expression of HMGA1 by Quantitative Real-Time PCR

RNA from 13/37 matched biopsies of HCC and their cirrhotic tissue previously subjected to transcriptomic profiling [28] was available and subjected to qRT-PCR analysis (Supplementary Methods).

Immunohistochemistry

Immunohistochemical staining of HMGA1 was assessed on a TMA of an independent cohort of 192 HCCs, 108 cirrhotic tissues and 79 normal liver samples, as previously described [29,30]. Follow-up information was available for 100/192 patients with HCC. HMGA1 antibody was raised against a synthetic peptide as previously reported [19]. Staining was performed as described previously [19,25] (Supplementary Methods). Samples with ≥5% HMGA1-positive cells were considered HMGA1-positive [19,25]. Staining was independently scored by three pathologists (DB, FT and LT).

Statistical Analysis

Statistical analyses for categorical and non-categorical variables were performed using Chi-Square/Fisher's Exact and Mann–Whitney U/Student's *t* tests. Analysis of the variance was performed using the ANOVA test. Correlation was assessed using Spearman's rank correlation. Survival analyses were performed using the Kaplan–Meier method and the log rank test. All tests were two-sided. *P*-values <0.05 were considered statistically significant. All analyses were performed using Graphpad Prism 6.0 (Graphpad Software, Inc., La Jolla, *CA*) or SPSS v.20 (Endicott, New York, NY).

Cell Lines

Four liver cancer cell lines (PCL5, HEPG2, SNU449 and SNU182) were used for *in vitro* experiments. All cell lines were negative for mycoplasma infection using the Universal Mycoplasma Detection kit (ATCC, Manassas, VA). Culture conditions are described in Supplementary Methods.

Vector Construction, Transfections of Mammalian Cells and Analysis of Transgene Expression

For overexpression, the pCDNA3.1-HMGA1 and the empty control vectors were constructed as previously described [10]. For down-regulation, the hairpin RNA interference plasmid for human HMGA1 (pLKO.1-HMGA1, TRCN0000018949) and the scramble control pLKO.1-Puro plasmid (SHC002) were purchased from Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO). The expression of HMGA1 in stable clones was evaluated by western blot (Supplementary Methods).

Proliferation Assay

Proliferation assays were performed using the xCELLigence Real-Time Cell Analysis (RTCA, ACEA Biosciences, San Diego, CA, USA) system. Cell index values were calculated and normalized by the RTCA Software Package v.1.2 (Supplementary Methods). Numerical data were expressed as mean ± standard deviation. Growth curves were analyzed using multiple t-tests, corrected for multiple comparisons by the Holm-Šídák method (alpha: 0.05) using GraphPad Prism 6.0 (Graphpad Software, Inc.).

Transwell Migration Assay

The transwell migration assay was used to assess the chemotactic and migration capacity of cells. Cells were seeded in the upper part of the transwell (8 µm pore membranes) of the 24-well plate and higher serum content was placed in the lower compartment to attract cells to migrate through the membrane. Cells that passed through the membrane were fixed on the membrane using methanol (Supplementary Methods). Fixed cells were stained with crystal violet and the number of migrated cells was determined by the Benchmark Plus microplate spectrophotometer (Bio-Rad, Hercules, CA, USA).

Soft Agar Colony Formation Assay

To assess cellular anchorage-independent growth *in vitro*, the soft agar colony formation assay, a stringent method for the detection of

the tumorigenic potential was used for the pCDNA3.1-HMGA1 and the pLKO.1-HMGA1 transformed cells (Supplementary Methods). Statistical analyses of the number and size of the colonies were performed with GraphPad Prism 6.0 (Graphpad Software, Inc.) using the Student's *t* test with Welch correction.

Results

HMGA1 mRNA is Frequently Up-Regulated in HCC

To determine the expression level of *HMGA1* in HCC, we re-analyzed a published gene expression microarray dataset of 59 HCC biopsies, 37 of which were matched with their respective non-tumoral cirrhotic tissues and 5 normal liver donor samples (GSE64041) [28]. Notably, none of the patients involved in the study received any therapeutic anti-cancer treatment at the time of biopsy.

Our analysis showed a monotonic increase in *HMGA1* mRNA expression from normal liver tissues to cirrhotic tissues to HCCs ($P < .001$; ANOVA test; Figure 1*A*), with increased expression in HCCs compared to cirrhotic tissues and normal liver tissues (both $P < .0001$, Mann-Whitney *U* tests) and in cirrhotic tissues compared to normal liver tissues ($P = .026$, Mann-Whitney U test; Figure 1*A*). Moreover, *HMGA1* expression levels were higher in Edmondson grades III/IV than in grade II tumors ($P < .0001$, Mann-Whitney *U* test; Figure 1*B*) and in HCCs with metastasis (regional lymph node invasion and/or distant organ involvement) than those without ($P = .0084$, Mann-Whitney *U* test; Figure 1*C*). When stratified by underlying virus or alcohol background, we found no difference in *HMGA1* expression levels, except that HCCs on a combination of virus and alcohol background showed higher expression than HCCs on an alcohol background ($P = .02$, Mann-Whitney *U* test; Figure 1*D*).

To confirm the array-derived data, we evaluated the *HMGA1* mRNA expression levels in 13 of the 37 paired HCCs and cirrhotic tissues for which RNA was available. Consistent with the array-derived data, we found that *HMGA1* expression was higher in HCCs than in cirrhotic tissues ($P = .0119$; paired Mann-Whitney *U* test; Supplementary Figure 1*A*). Furthermore, *HMGA1* expression levels from array-derived and qRT-PCR data were highly correlated ($r = 0.67$, Spearman correlation, Supplementary Figure 1*B*). Altogether, these results demonstrate that *HMGA1* mRNA is overexpressed in HCCs than their matched cirrhotic tissues and normal liver tissues.

HMGA1 Protein Expression is Associated With Disease Progression and Poorer Survival in HCC

To corroborate our mRNA expression-derived results on the protein level, we evaluated HMGA1 on a TMA of an independent cohort of 192 HCCs, 108 cirrhotic tissues and 79 normal liver samples [29,30]. Consistent with the mRNA expression analysis, the percentage of HMGA1-positive cells increased monotonically

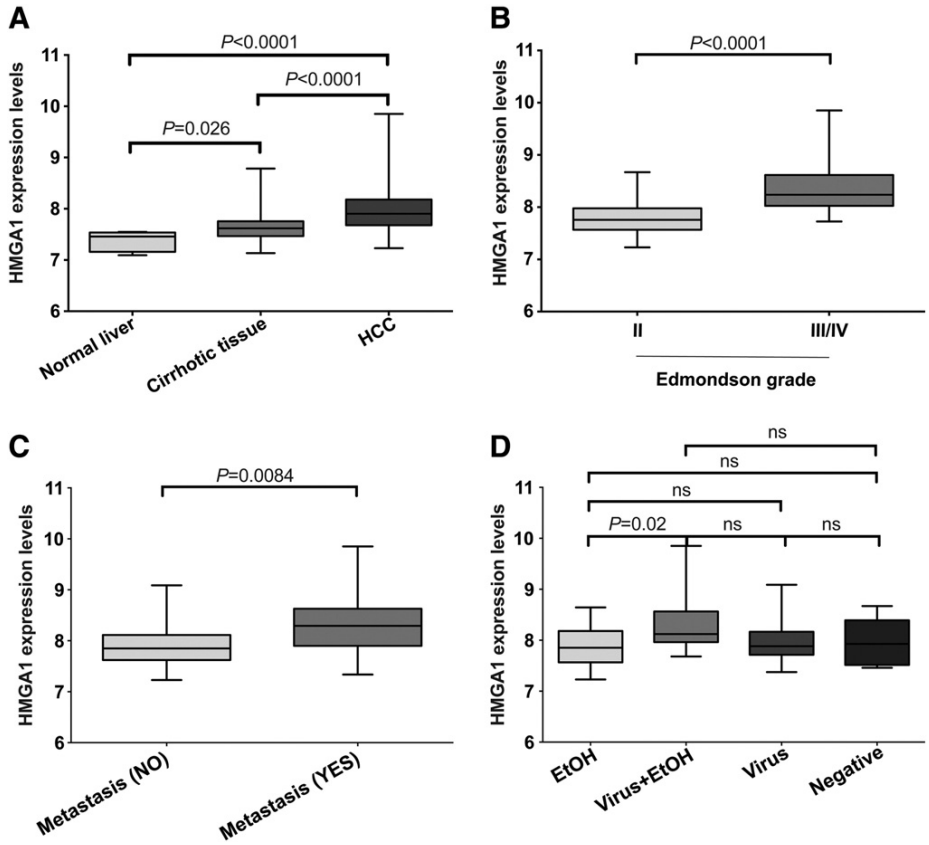


Figure 1. *HMGA1 mRNA* expression level using gene expression microarrays. Boxplots show *HMGA1* expression (A) in HCC area, corresponding non-tumoral area (cirrhotic tissue) and normal liver samples, (B) in moderately and poorly differentiated HCCs, (C) in HCCs associated with and not associated with metastasis and (D) in HCCs stratified according to the etiology. Statistical comparisons were performed using Mann-Whitney *U* tests. $P < .05$ was considered statistically significant. EtOH: alcohol-related. ns: not significant.

through the progression stages from normal liver to cirrhotic tissue to HCC ($P < .0001$; ANOVA test; Figure 2*A*) and were significantly higher in HCCs and in cirrhotic tissues than in normal liver samples (both $P < .0001$, Mann–Whitney U tests; Figs. 2a and 3). HMGA1 expression was also increased in Edmondson grades III/IV and II HCCs compared to grade I HCCs ($P = .037$ and $P = .018$, respectively, Mann–Whitney U tests; Figure 2*B*) with a monotonic increase from grade I to grades III/IV ($P = .026$; ANOVA test). Overall, 51%, 42% and 5% of HCCs, cirrhotic tissues and normal liver samples expressed HMGA1, respectively ($P < .0001$, Chi-squared test; Table 1) and 74%, 52% and 43% of Edmonson grades III/IV, II and I HCCs expressed HMGA1, respectively ($P = .031$, chi-squared test; Table 2). There was no association between HMGA1 positivity and other clinicopathologic parameters, except the male gender (Table 2).

We further explored whether HMGA1 positivity was associated with clinical progression and outcome in HCCs. Of the 192 HCCs, clinical follow-up was available for 100 patients. We found that patients with HMGA1-positive HCCs were associated with earlier disease progression and worse overall survival ($P = .01$ and $P = .025$, respectively, log rank tests, Figure 2, *C* and *D*).

HMGA1 Overexpression Promote Tumor Growth and Migration in In Vitro Models

To define whether HMGA1 would have oncogenic properties in *in vitro* models of HCC, we tested the effect of its overexpression using cDNA constructs in stable cultures of PCL5 and HEPG2 with low HMGA1 endogenous expression and of its silencing in stable cultures of SNU449 and SNU182 with high HMGA1 endogenous expression (Supplementary Figure 2). Forced expression of HMGA1 in PCL5 and HEPG2 led to increased cell growth (both $P < .01$, Figure 4, *A* and *C*), while HMGA1 silencing in SNU449 and SNU182 showed reduced cell growth (both $P < .01$, Figure 4, *E* and *G*) suggesting that HMGA1 expression promotes liver cancer cell growth. Using a transwell migration assay, forced expression of HMGA1 in PCL5 and HEPG2 resulted in increased cell migration (both $P < .05$, Figure 4, *B* and *D*) and conversely, HMGA1 silencing in SNU449 and SNU182 led to decreased cell migration (both $P < .01$, Figure 4, *F* and *H*).

We next investigated whether modulating HMGA1 expression would alter the transformation capacity of liver cancer cells. By assessing the impact of HMGA1 on anchorage-independent growth on soft agar, we found that PCL5 and HEPG2 cells stably overexpressing HMGA1 led to increased number and size of colonies

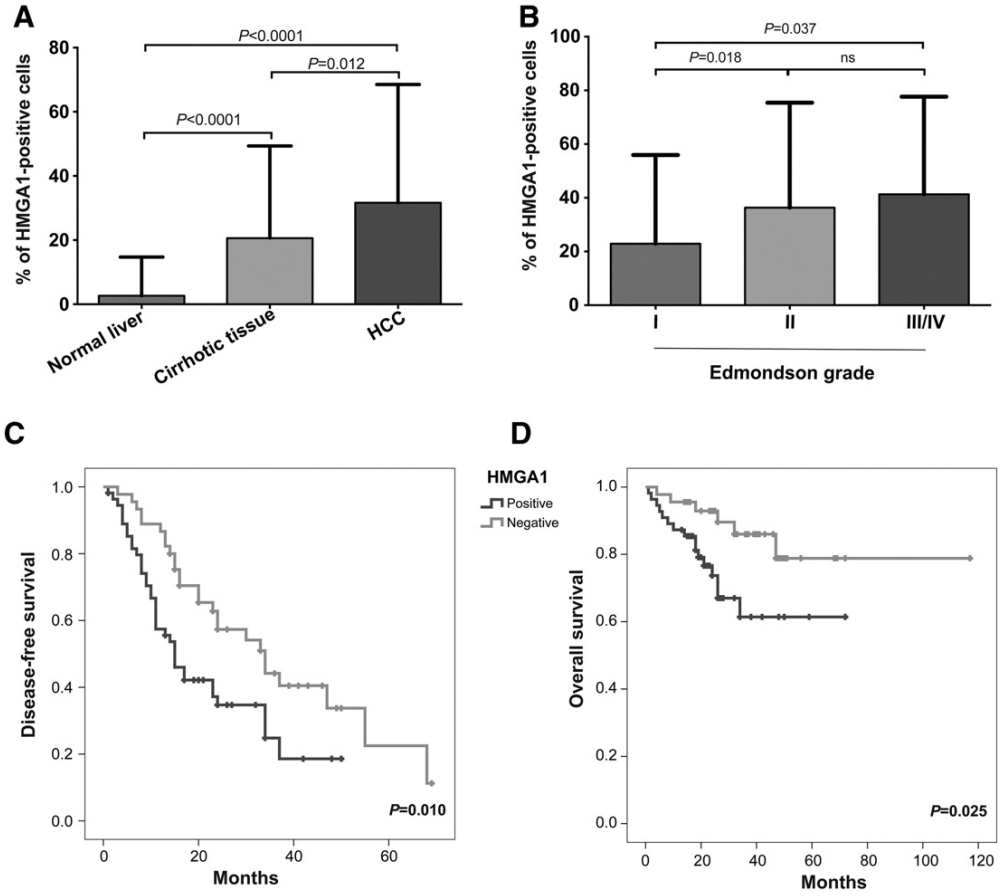


Figure 2. HMGA1 protein expression level using TMA. Boxplots show HMGA1 expression (A) in HCCs, cirrhotic tissues and normal liver samples, and (B) in well, moderately and poorly differentiated HCCs. (C) Disease-free and (D) overall survival of patients with HCCs that expressed and did not express HMGA1 using the Kaplan-Meier method. Statistical comparisons were performed using (A and B) Mann-Whitney *U* tests and (C and D) log-rank tests. $P < .05$ was considered statistically significant. ns: not significant.

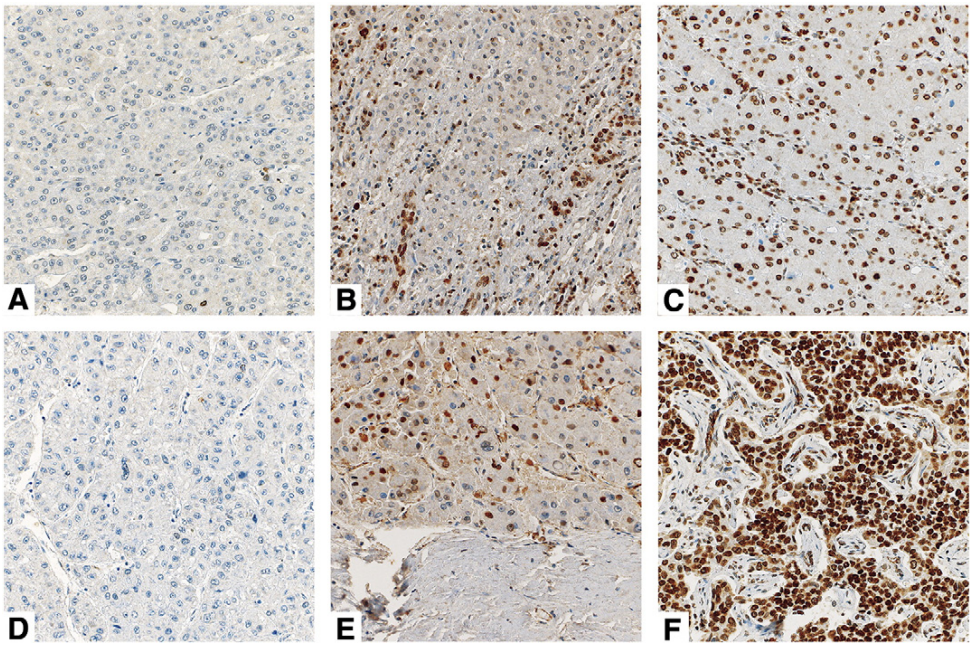


Figure 3. Representative micrographs of HMGA1 protein staining. Representative micrographs of (A) negative, (B) moderate and (C) high HMGA1 expression in cirrhotic tissues, and of (D) negative, (E) moderate and (F) high HMGA1 expression in HCC. All the micrographs were taken at 20×.

(both at $P < .001$, Student t-tests with Welch correction; Figure 5, *A* and *B*). By contrast, down-regulating HMGA1 in stable clones of SNU449 and SNU182 led to decreased number and size of colonies (both at $P < .001$, Student *t* tests with Welch correction; Figure 5, *C* and *D*).

Taken together our results suggest that, in line with previous reports in other cancer entities, HMGA1 overexpression plays a pivotal role in cell viability enhancing cell growth and migration and induces transformation by anchorage-independent growth in liver cancer cell lines.

Discussion

In this study, we demonstrated in two independent cohorts that HMGA1 levels monotonically increased through the stages of progression from normal liver to cirrhosis to HCC, both at the mRNA and the protein levels and that HMGA1 protein expression was associated with poor disease-free and overall survival. Our hypothesis that HMGA1 is a driver of progression of HCC is supported by our functional evidence that HMGA1 promoted cell growth, migration and transformation in liver cancer cell lines. These results provide evidence that HMGA1 confers a neoplastic advantage to liver cancer cell lines.

Of particular interest is that HMGA1 is expressed in 42% of cirrhotic tissues. HMGA1 expression has been observed in other preneoplastic conditions, including colon adenomas, pancreatic intraepithelial neoplasias and breast hyperplasia [16,18,19]. In fact, a similar pattern of monotonically increasing HMGA1 expression through progression has been found in normal colon epithelium, colon adenoma and colorectal carcinoma [16] and in normal pancreatic tissue, pancreatic intraepithelial neoplasias and invasive ductal adenocarcinoma of the pancreas [19]. Crucially, the substantial proportion of cirrhotic tissues showing high HMGA1 expression suggests that although HMGA1 is a driver of progression, it is not a specific biomarker for HCC.

Consistent with the reported survival differences between HCCs that did or did not express *HMGA1* mRNA [27], we found that 51% of HCCs were HMGA1-positive by IHC and that HMGA1 protein

Table 1. Analysis of HMGA1 Expression by Immunohistochemistry

	HMGA1-negative	HMGA1-positive	% HMGA1-positive	*P*-value
Normal Liver (n = 79)	75	4	5%	
Cirrhotic tissue (n = 108)	63	45	42%	<.0001
HCC (n = 192)	95	97	51%	

Statistical Comparison was Performed Using Chi-Squared Test

Table 2. Analysis of HMGA1 Expression by Immunohistochemistry in 192 HCCs

Clinicopathologic information		HMGA1-negative	HMGA1-positive	% HMGA1-positive	*P*-value
Gender	Female	27	14	34%	.021
	Male	67	84	56%	
Tumor Stage	I/II	58	63	52%	.869
	III/IV	24	28	54%	
N stage	0	83	88	51%	.378
	1	4	8	67%	
M stage	0	71	85	54%	.214
	1	16	11	41%	
Multifocality	No	41	43	51%	1
	Yes	49	52	51%	
Vascular Invasion	No	53	53	50%	.526
	Yes	28	35	56%	
Etiology	EtOH	17	19	53%	.788
	Virus	43	54	56%	
	Other	4	3	43%	
Edmondson grade	I	44	33	43%	.031
	II	42	45	52%	
	III/IV	6	17	74%	

Statistical Comparisons were Performed Using Chi-Squared Tests

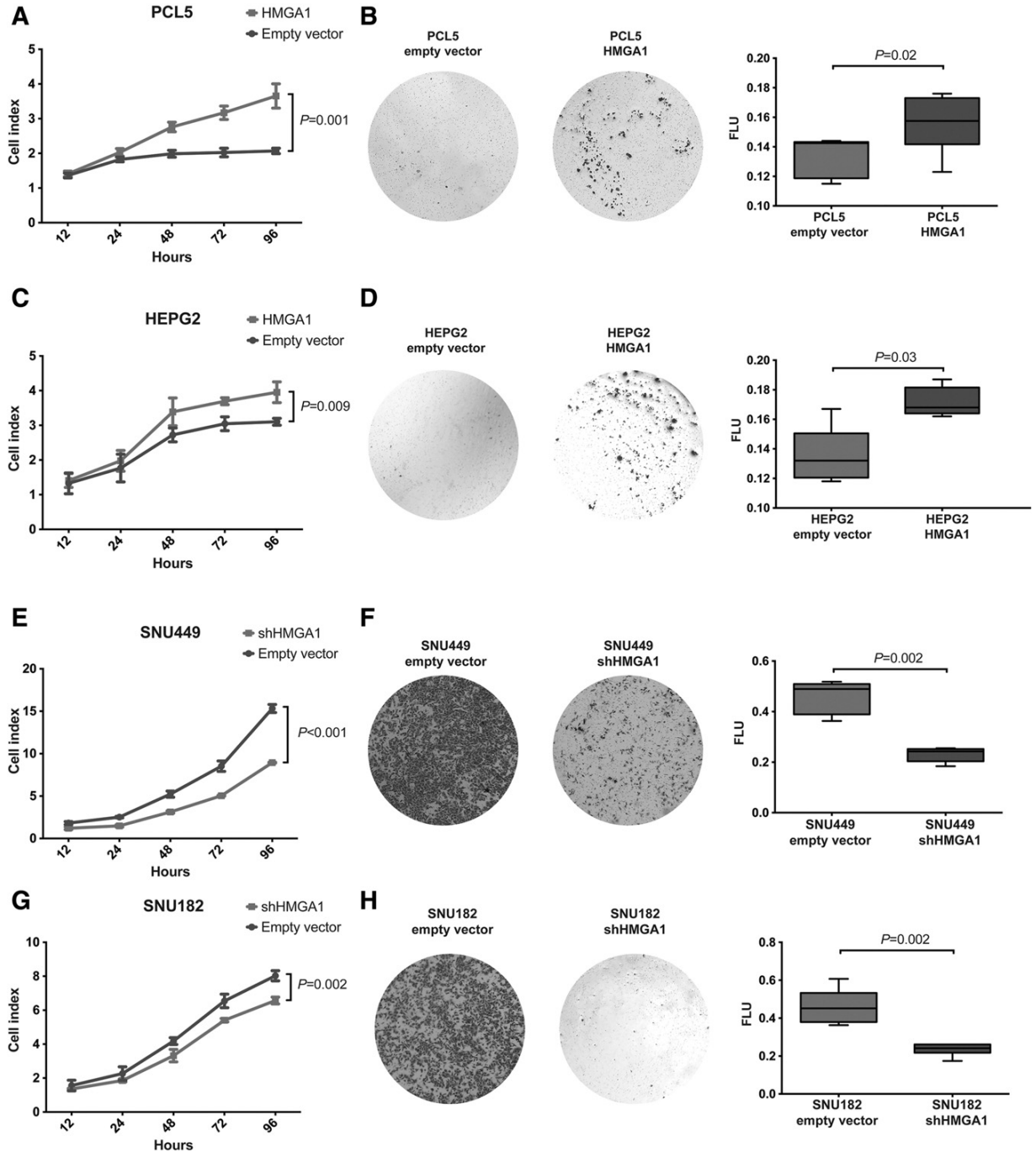


Figure 4. Impact of HMGA1 on cell growth and migration in *in vitro* models. Effect of overexpression of HMGA1 in (A) PCL5 and (C) HEPG2 and down-regulation of HMGA1 in (E) SNU449 and (G) SNU182 on cell growth compared to cells transfected with empty vector control. Effect of overexpression of HMGA1 in (B) PCL5 and (D) HEPG2 and down-regulation of HMGA1 in (F) SNU449 and (h) SNU182 on cell migration using Transwell assays compared to cells transfected with empty vector. Quantification was performed using a spectrophotometer (FLU: Fluorescence spectroscopy of dyes). All experiments were performed in triplicates. Error bars, SD of the mean. Statistical comparisons were performed using (A, C, E, G) Holm-Šídák-corrected multiple *t* tests and (B, D, F, H) Mann-Whitney *U* tests. $P < .05$ was considered statistically significant.

expression confers worse prognosis in HCC. While the previous study reported that 13% of HCCs were HMGA1-positive by IHC [27], the higher frequency of HMGA1 positivity we found may be attributed to a much larger cohort and a different antibody for immunodetection. In terms of prognosis, we found that HMGA1 expression conferred worse prognosis, similar to the association of shorter survival in patients with pancreatic ductal carcinoma that showed strong immunoreactivity [20], and to the association of shorter disease-free and overall survival, as well as an increased risk of distant metastases, in patients with HMGA1-overexpressing uveal melanoma [31]. Here we demonstrated

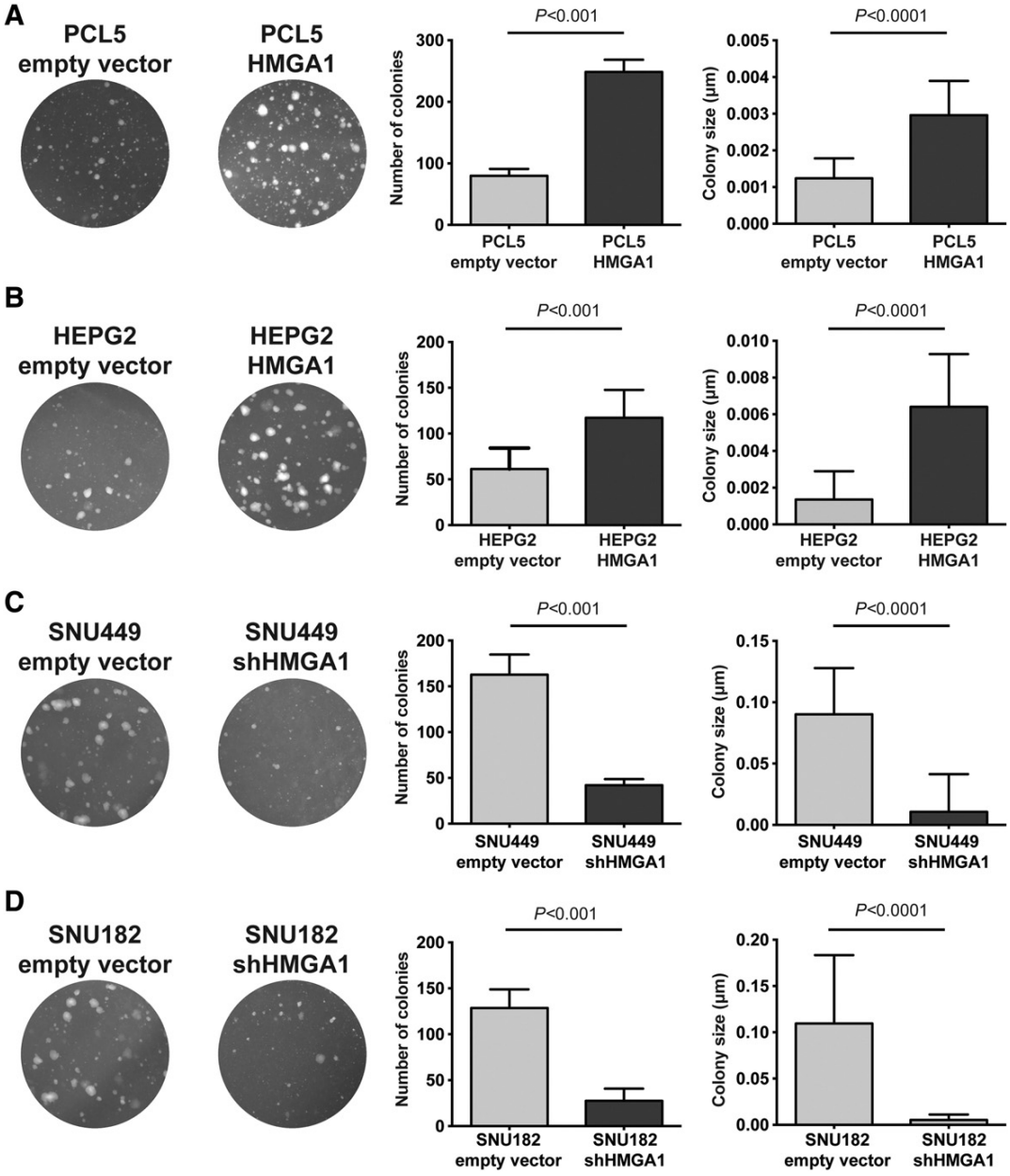


Figure 5. Impact of HMGA1 on cell transformation in *in vitro* models. Effect of overexpression of HMGA1 in (A) PCL5 and (B) HEPG2 and down-regulation of HMGA1 in (C) SNU449 and (D) SNU182 on anchorage-independent growth. Quantification was performed by defining the number and size of colonies. All experiments were performed in triplicates. Error bars, SD of the mean. Statistical comparisons were performed using unpaired t-tests with Welch correction. $P < .05$ was considered statistically significant.

that HMGA1 is up-regulated in a substantial proportion of HCC and its expression is associated with poor prognosis.

Previous studies showed that HMGA1 triggers oncogenic transformation in cultured cells [32] and is associated with aggressive cancer subtypes in animal models in several tumor types [20,33–35]. For example, in breast cancer cells, *HMGA1* overexpression directly activates genes involved in tumor cell migration and invasion [6] and induces epithelial-to-mesenchymal transition [36]. Indeed, we demonstrated in multiple liver cancer cell line models that the HMGA1 increases tumor cell growth and migration and that HMGA1 increased transformation potential in liver cancer cells. Nonetheless, these results are consistent with our and others' [27] observation that HMGA1 expression was associated with Edmondson grade and support the role of HMGA1 as a driver of progression.

This study has limitations. We studied HMGA1 protein expression on TMA punches rather than whole sections, thus the observed expression may not be representative of the individual tissue samples. Despite this, given the large cohort, we expect that the results to be representative on the cohort level. Secondly, the HCCs in the TMA cohort were from resected materials rather than untreated biopsies, thus the HMGA1 levels may have been altered as a result of the surgical procedures and the long post-surgical hypoxia derived for

the lack of blood perfusion. However, the analysis of the expression microarrays of untreated liver biopsies is in agreement with the findings at the protein level and it should be emphasized that the liver biopsies, unlike resected materials, had never been subjected to HCC-tailored therapies and are thus the most representative of the natural biology of HCC. In conclusion, our findings demonstrated a functional role for HMGA1 in the progression of HCC. Given multi-faceted functions of HMGA1, further characterization of its function in liver biology will provide novel insights into its mechanisms of action in driving disease progression.

Conflict of Interest

The authors have no conflicts of interest to declare.

Role of the Funding Source

Funding bodies had no role in the design of the study, collection, analysis and interpretation of the data or the writing of the manuscript.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at doi:10.1016/j.neo.2016.10.002.

References

[1] Bianchi ME and Beltrame M (2000). Upwardly mobile proteins. Workshop: the role of HMG proteins in chromatin structure, gene expression and neoplasia. *EMBO Rep* **1**(2), 109–114.

[2] Reeves R (2000). Structure and function of the HMGI(Y) family of architectural transcription factors. *Environ Health Perspect* **108**(Suppl. 5), 803–809.

[3] Reeves R (2001). Molecular biology of HMGA proteins: hubs of nuclear function. *Gene* **277**(1–2), 63–81.

[4] Thomas JO (2001). HMG1 and 2: architectural DNA-binding proteins. *Biochem Soc Trans* **29**(Pt 4), 395–401.

[5] Reeves R and Adair JE (2005). Role of high mobility group (HMG) chromatin proteins in DNA repair. *DNA Repair (Amst)* **4**(8), 926–938.

[6] Fusco A and Fedele M (2007). Roles of HMGA proteins in cancer. *Nat Rev Cancer* **7**(12), 899–910.

[7] Fedele M and Fusco A (2010). HMGA and cancer. *Biochim Biophys Acta* **1799**(1–2), 48–54.

[8] Zhou G, Chen J, Lee S, Clark T, Rowley JD, and Wang SM (2001). The pattern of gene expression in human CD34(+) stem/progenitor cells. *Proc Natl Acad Sci U S A* **98**(24), 13966–13971.

[9] Giancotti V, Berlingieri MT, DiFiore PP, Fusco A, Vecchio G, and Crane-Robinson C (1985). Changes in nuclear proteins on transformation of rat epithelial thyroid cells by a murine sarcoma retrovirus. *Cancer Res* **45**(12 Pt 1), 6051–6057.

[10] Puca F, Colamaio M, Federico A, Gemei M, Tosti N, and Bastos AU, et al (2014). HMGA1 silencing restores normal stem cell characteristics in colon cancer stem cells by increasing p53 levels. *Oncotarget* **5**(10), 3234–3245.

[11] Shah SN, Kerr C, Cope L, Zambidis E, Liu C, and Hillion J, et al (2012). HMGA1 reprograms somatic cells into pluripotent stem cells by inducing stem cell transcriptional networks. *PLoS One* **7**(11), e48533.

[12] Hillion J, Wood LJ, Mukherjee M, Bhattacharya R, Di Cello F, and Kowalski J, et al (2009). Upregulation of MMP-2 by HMGA1 promotes transformation in undifferentiated, large-cell lung cancer. *Mol Cancer Res* **7**(11), 1803–1812.

[13] Resar LM (2010). The high mobility group A1 gene: transforming inflammatory signals into cancer? *Cancer Res* **70**(2), 436–439.

[14] Takaha N, Resar LM, Vindivich D, and Coffey DS (2004). High mobility group protein HMGI(Y) enhances tumor cell growth, invasion, and matrix metalloproteinase-2 expression in prostate cancer cells. *Prostate* **60**(2), 160–167.

[15] Tesfaye A, Di Cello F, Hillion J, Ronnett BM, Elbahloul O, and Ashfaq R, et al (2007). The high-mobility group A1 gene up-regulates cyclooxygenase 2 expression in uterine tumorigenesis. *Cancer Res* **67**(9), 3998–4004.

[16] Abe N, Watanabe T, Sugiyama M, Uchimura H, Chiappetta G, and Fusco A, et al (1999). Determination of high mobility group I(Y) expression level in colorectal neoplasias: a potential diagnostic marker. *Cancer Res* **59**(6), 1169–1174.

[17] Chiappetta G, Manfioletti G, Pentimalli F, Abe N, Di Bonito M, and Vento MT, et al (2001). High mobility group HMGI(Y) protein expression in human colorectal hyperplastic and neoplastic diseases. *Int J Cancer* **91**(2), 147–151.

[18] Chiappetta G, Botti G, Monaco M, Pasquinelli R, Pentimalli F, and Di Bonito M, et al (2004). HMGA1 protein overexpression in human breast carcinomas: correlation with ErbB2 expression. *Clin Cancer Res* **10**(22), 7637–7644.

[19] Piscuoglio S, Zlobec I, Pallante P, Sepe R, Esposito F, and Zimmermann A, et al (2012). HMGA1 and HMGA2 protein expression correlates with advanced tumour grade and lymph node metastasis in pancreatic adenocarcinoma. *Histopathology* **60**(3), 397–404.

[20] Hristov AC, Cope L, Di Cello F, Reyes MD, Singh M, and Hillion JA, et al (2010). HMGA1 correlates with advanced tumor grade and decreased survival in pancreatic ductal adenocarcinoma. *Mod Pathol* **23**(1), 98–104.

[21] Masciullo V, Baldassarre G, Pentimalli F, Berlingieri MT, Boccia A, and Chiappetta G, et al (2003). HMGA1 protein over-expression is a frequent feature of epithelial ovarian carcinomas. *Carcinogenesis* **24**(7), 1191–1198.

[22] Zhang Z, Wang Q, Chen F, and Liu J (2015). Elevated expression of HMGA1 correlates with the malignant status and prognosis of non-small cell lung cancer. *Tumour Biol* **36**(2), 1213–1219.

[23] Kettunen E, Anttila S, Seppanen JK, Karjalainen A, Edgren H, and Lindstrom I, et al (2004). Differentially expressed genes in nonsmall cell lung cancer: expression profiling of cancer-related genes in squamous cell lung cancer. *Cancer Genet Cytogenet* **149**(2), 98–106.

[24] Franco R, Esposito F, Fedele M, Liguori G, Pierantoni GM, and Botti G, et al (2008). Detection of high-mobility group proteins A1 and A2 represents a valid diagnostic marker in post-pubertal testicular germ cell tumours. *J Pathol* **214**(1), 58–64.

[25] Sepe R, Piscuoglio S, Quintavalle C, Perrina V, Quagliata L, and Formisano U, et al (2015). HMGA1 overexpression is associated with a particular subset of human breast carcinomas. *J Clin Pathol*.

[26] Guichard C, Amaddeo G, Imbeaud S, Ladeiro Y, Pelletier L, and Maad IB, et al (2012). Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat Genet* **44**(6), 694–698.

[27] Chang ZG, Yang LY, Wang W, Peng JX, Huang GW, and Tao YM, et al (2005). Determination of high mobility group A1 (HMGA1) expression in hepatocellular carcinoma: a potential prognostic marker. *Dig Dis Sci* **50**(10), 1764–1770.

[28] Makowska Z, Boldanova T, Adametz D, Quagliata L, Vogt JE, and Dill MT, et al (2016). Gene expression analysis of biopsy samples reveals critical limitations of transcriptome-based molecular classifications of hepatocellular carcinoma. *J Pathol Clin Res* **2**, 80–92.

[29] Quagliata L, Andreozzi M, Kovac M, Tornillo L, Makowska Z, and Moretti F, et al (2014). SH2D4A is frequently downregulated in hepatocellular carcinoma and cirrhotic nodules. *Eur J Cancer* **50**(4), 731–738.

[30] Baumhoer D, Tornillo L, Stadlmann S, Roncalli M, Diamantis EK, and Terracciano LM (2008). Glypican 3 expression in human nonneoplastic, preneoplastic, and neoplastic tissues: a tissue microarray analysis of 4,387 tissue samples. *Am J Clin Pathol* **129**(6), 899–906.

[31] Qu Y, Wang Y, Ma J, Zhang Y, Meng N, and Li H, et al (2013). Overexpression of high mobility group A1 protein in human uveal melanomas: implication for prognosis. *PLoS One* **8**(7), e68724.

[32] Wood LJ, Maher JF, Bunton TE, and Resar LM (2000). The oncogenic properties of the HMG-I gene family. *Cancer Res* **60**(15), 4256–4261.

[33] Tanaka M, Itoh T, Tanimizu N, and Miyajima A (2011). Liver stem/progenitor cells: their characteristics and regulatory mechanisms. *J Biochem* **149**(3), 231–239.

[34] Sarhadi VK, Wikman H, Salmenkivi K, Kuosma E, Sioris T, and Salo J, et al (2006). Increased expression of high mobility group A proteins in lung cancer. *J Pathol* **209**(2), 206–212.

[35] Flohr AM, Rogalla P, Bonk U, Puettmann B, Buerger H, and Gohla G, et al (2003). High mobility group protein HMGA1 expression in breast cancer reveals a positive correlation with tumour grade. *Histol Histopathol* **18**(4), 999–1004.

[36] Dolde CE, Mukherjee M, Cho C, and Resar LM (2002). HMG-I/Y in human breast cancer cell lines. *Breast Cancer Res Treat* **71**(3), 181–191.