LIVER PROGENITOR CELL MARKERS CORRELATE WITH LIVER DAMAGE AND PREDICT SHORT-TERM MORTALITY IN PATIENTS WITH ALCOHOLIC HEPATITIS

Pau Sancho-Bru[1], José Altamirano[1], Daniel Rodrigo-Torres[1], Mar Coll[1], Cristina Millán[1], Juan José Lozano[1], Rosa Miquel[2], Vicente Arroyo[1], Juan Caballería[1], Pere Ginès[1], Ramon Bataller[1].

[1]*Liver Unit, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Barcelona, Spain.*

[2] *Pathology Unit, Hospital Clínic, Barcelona, Spain.*

Short title: Liver Progenitor Cells in Alcoholic Hepatitis

Keywords: Liver regeneration, liver progenitor cells, stem cells, alcohol-induced liver disease, liver disease

Corresponding author:

Pau Sancho-Bru, PhD

Institut d'Investigacions Biomèdiques August Pi i Suyner (IDIBAPS)

C/ Roselló, 153, Planta 3,

08036 Barcelona, Spain

psancho@clinic.ub.es

Tel. +34 932275400 Ext. 4535

Fax. +34 93 3129406

ABREVIATIONS

Alcoholic hepatitis (AH)

Liver progenitor cells (LPC)

Epithelial cell adhesion molecule (EpCAM)

Prominin-1 (PROM1)

Keratin (KRT)

Classification and regression tree (CART)

Hepatitis C virus (HCV)

Grant support: This work was supported by grants from the Instituto de Salud Carlos III (FIS PI080237, FIS PS09/01164 and FIS PI080126 to Dr Bataller, Dr Caballería and Dr Ginès, respectively). Dr Sancho-Bru received a grant from Ministerio de Ciencia e Innovación, Juan de la Cierva (JCI-2009-03849) and Instituto de Salud Carlos III, Miguel Servet (CP11/00071). D. Rodrigo-Torres received a grant from the Ministerio de Educación, FPU program. J. Altamirano received a grant from Fundación Banco Bilbao Vizcaya Argentaria.

ABSTRACT

Alcoholic Hepatitis (AH) is a severe condition developed in patients with underlying alcoholic liver disease. Ductular reaction has been associated with chronic alcohol consumption but there is no information regarding the extent of liver progenitor cell (LPC) proliferation in AH. The aim of this study was to investigate LPC markers in AH, and its correlation with disease severity. Fifty-nine patients with clinical and histological diagnosis of AH were included in the study. LPC markers were assessed by real time PCR and immunohistochemistry. Standard logistic regression analysis and classification and regression trees (*CART*) analysis were used for statistical analysis. A microarray analysis showed an up-regulation of LPC markers in patients with AH. Real time PCR demonstrated that *epithelial cell adhesion molecule (EpCAM)*, *Prominin-1*, and *Keratin7* were significantly increased in patients with AH compared to normal livers (p≤0,01), chronic hepatitis C (p≤0,01), and HCV-induced cirrhosis (p≤0,01). Immunohistochemistry scores generated for Keratin7 and EpCAM demonstrated a good correlation with gene expression. *Keratin7* gene expression correlated with liver failure as assessed by MELD score (r=0.41, p =0.006) and Maddrey's discriminant function (r=0.43, p=0.004). Moreover, *Keratin7* (HR1.14, p=0.004) and *Prominin-1* (HR1.14, p=0.002), but not *EpCAM* (HR1.16, p=0.06), were identified as independent predictors of 90-days mortality. CART analysis generated an algorithm based-on the combination of *Keratin7* and *EpCAM* gene expression that stratified three groups of patients with high, intermediate and low short-term mortality (89%, 33% and 6% respectively; AUROC 0.73, CI 95%: 0.60-0.87). Keratin7 expression provided additional discrimination potential to ABIC score. *Conclusion*: LPC markers correlate positively with severity of liver disease and short-term mortality in AH patients. This study suggests that LPC proliferation may be an important feature of AH pathophysiology.

Alcoholic liver disease is a major cause of end stage liver disease. It ranges from fatty liver disease to steatohepatitis, cirrhosis and eventually hepatocellular carcinoma (1, 2). Alcoholic hepatitis (AH) is an acute event on a chronic alcoholic liver disease that develops in up to 20% of patients with a heavy alcohol intake (3). In its severe forms, AH leads to liver failure and high short-term mortality rate. AH is characterized by an important hepatocellular damage, megamitochondria, hepatocyte arrest, inflammatory response, and rapid progression of fibrosis (4). We have recently developed the ABIC score, which allows a prognostic stratification based on analytical parameters: age, bilirubin, INR and creatinine into patients with low, intermediate and high risk of death at 90 days and one year (5). The pathogenesis of AH is poorly understood and current therapies are not fully effective. Thus, there is a clear need to better understand the pathogenesis of AH in order to identify new targets for therapy and develop new therapeutic strategies for these patients.

Liver damage from any etiology induces mature hepatocyte to proliferate in order to replace the damaged tissue, allowing the recovery of liver function without any involvement of liver progenitor cells (LPC) (6, 7). However, when there is either a massive liver injury or a chronic liver damage that compromise the proliferative capacity of hepatocytes, progenitor cells within the Canal of Hering start to proliferate giving rise to what is known as ductular reaction (6-8). Little is known about the real contribution of LPC in the recovery of liver function in liver diseases. Immunohistochemistry studies performed in damaged human livers have shown that ductular reaction comprises a heterogenic population of proliferating cells, ranging from cells expressing stem cells markers with an immature phenotype, to more committed cells with an intermediate phenotype, expressing progenitor cell markers but also characteristics of both immature hepatocytes and cholangiocytes (9-12). KRT7 is a well-known marker of ductular reaction in liver disease and it is typically expressed in LPC but also in intermediate hepatobiliary cells (13, 14). PROM1 has been described in progenitor cells from the liver and other organs and is considered a marker for hepatic cancer stem cells (15, 16). EpCAM is

expressed in LPC and it has been used to isolate progenitor cells from human samples (17, 18). Moreover, EpCAM is also expressed in newly generated hepatocytes derived from progenitor cells in the regenerating liver (19). *In vitro* studies with isolated LPC from human liver samples have demonstrated that LPC can differentiate into hepatocytes and cholangiocytes exerting functions of mature cells (17, 20, 21). However, there are no functional studies that investigate the dynamics, progression and outcome of progenitor cell expansion during the course of liver disease including AH.

Ductular reaction is known to be present in most chronic liver diseases but it seems especially important in acute-on-chronic events (22, 23). Moreover, factors such as fibrosis and inflammation are known to promote LPC expansion. Advanced alcoholic liver disease is characterized by an important ductular reaction due to the underling liver injury but also to the effect of alcohol in promoting LPC proliferation (24, 25). Moreover, the extent of ductular reaction correlates with the severity of liver disease (23, 24). AH is characterized by an impairment in cell proliferation together with a massive liver damage and inflammatory response, which are key factors influencing the progression of progenitor cells. However, there is no information on the extent of ductular reaction in AH and its correlation with the severity of liver disease. In the present paper we describe that in AH there is an important ductular reaction and extensive proliferation of progenitor cells. Moreover, progenitor cell markers correlate with liver disease and are good prognostic markers for short-term mortality in patients with AH.

METHODS

Patients and Interventions

This study has been performed with samples collected from a cohort of sixty-nine consecutive patients admitted to the Liver Unit of the Hospital Clinic, Barcelona, from 2007 to 2010, with clinical and analytical criteria of AH. Ten patients were excluded from the study, eight patients that did not fulfilled all diagnostic criteria of AH, and two patients with documented hepatocellular carcinoma. Finally, 59 patients with biopsy-proven AH were included. Liver biopsies were obtained by transjugular approach in all cases within 48 hours of admission and before starting treatment with corticosteroids if indicated. Patients were managed following international guidelines and the clinical protocols approved in the Liver Unit of the hospital (26-28). Healthy patients characteristics are depicted in supplementary figure 1. The study was approved by the Ethics Committee of the Hospital Clinic, and all patients included in this study gave written informed consent.

Microarray data normalization

Markers of LPCs upregulated in the AH microarray were identified and summarized. The microarray previously performed comprised data from 22 samples, 15 AH and 7 healthy samples (29). Information regarding microarray analysis can be found in Supplementary Material.

Hepatic gene expression analysis

Quantitative PCR was used to assess the expression of selected genes in liver human samples.

then amplified using Taqman Technology (Applied Biosystems) in a final PCR volume of 10 µl using a 7900 HT instrument (Applied Biosystems). The Assay-on-Demand probes and primers for the quantification of 18s, *EpCam, Prominin-1(PROM1)* and *Keratin (KRT)7* (Ref seq: Hs99999901_s1, Hs00158980_m1, Hs01009259_m1 and Hs00559840_m1) were obtained from Applied Biosystems. Results were normalized to 18s rRNA expression (housekeeping gene) and gene expression values were calculated based on the ΔΔCt method using a pool of liver RNA as internal reference. The results were expressed as $2^{(-\Delta\Delta Ct)}$.

Histological analysis and development of KRT7 and PROM-1 score

Three µm of paraffin-embeeded liver sections were incubated with mouse anti-human KRT7 (1:50, DakoCytomation, Glostrup, Denmark, M7018) for 1h at room temperature and monoclonal mouse anti-human Epithelial Antigen (EpCAM) (1:100, DakoCytomation, M0804) for 2h at room temperature. After washing in PBS, sections were incubated with secondary goat anti-mouse antibody conjugated to HRP (DAKO EnVision System-HRP, K4007) for 30 minutes at room temperature. 3,3' diaminobenzidine (DAB, Dako) was used as a chromogen, and sections were counterstained with hematoxilin. As negative controls, all specimens were incubated with primary antibody omission under identical conditions. Dual immunofluorescent staining was performed as serial stainings with EpCAM, KRT7 and the hepatocyte marker clone OCH1e5 (M7158) (DAKO). (See Supplementary Material). Semiquantitative assessment of inflammation and fibrosis was performed by RM on hematoxylin and eosin staining and tricromic staining respectively. Quantification of fibrosis was performed by morphometric quantification of tricromic stained area. (See Supplementary Material).

To assess the degree of expression of KRT7 and EpCAM, we developed a semi-quantitative score based on the nomenclature and analysis of ductular reaction previously described (9, 10). Staining for KRT7 was quantified according to 3 parameters punctuated from 0-3: Ductular structures, isolated cells within the parenchyma and Intermediate hepatobiliary Cells. Staining

for EpCAM was quantified according to 3 parameters punctuated from 0-3: ductular structures, interphase Cells and hepatobiliary cells

Statistical Analysis

Continuous variables were described as mean (95% confidence interval) or median (inter-quartile range). Categorical variables were described by means and percentages. Comparisons between groups were performed using the Student´s t test or Mann-Whitney U test, depending on variable distribution. Differences between categorical variables were assessed by the chi-square test or Fisher´s exact test, when necessary. Correlations between variables were evaluated using Spearman's *rho* or Pearson's *r*, when appropriate. The main end-point was death from any cause at 90 days. A two-sided *p* value of <0.05 was required for statistical significance. The potential role of LPC markers in the short-term prognosis of patients with AH was evaluated by a univariate analysis, multivariate logistic regression analysis. A classification and regression tree (CART) analysis was performed to evaluate the interaction of LPC markers with short-term prognosis of AH patients. (See Supplementary Material)

RESULTS

Patients Characteristics

The baseline clinical, demographical and biochemical characteristics of the study cohort are depicted in Table 1. The median alcohol consumption was 100 g/day (80-120). Twenty-four patients (41%) were treated with corticosteroids. During hospitalization, thirty patients (51%) developed at least one clinical complication. These complications included bacterial infections (36%), hepatic encephalopathy (21%), in-hospital renal dysfunction (20%) and gastrointestinal bleeding (7%). The overall 90-day mortality rate was 27% (16 out of 59 patients). The main causes of death at 90 day were sepsis (7 out of 16) and multiorgan failure (9 out of 16).

Expression of liver progenitor cell markers in patients with alcoholic hepatitis.

Previous studies have shown that alcoholic liver disease is characterized by an important ductular reaction and LPC expansion (24, 25). Thus, we have evaluated the hepatic expression of genes typically expressed in LPC in patients with AH and from healthy subjects. Microarray analysis demonstrated an up-regulation of genes related to LPC and ductular reaction in AH samples. These genes include KRT7, SOX9, EpCAM, KRT19, PROM1, CD44 and others. Although some of these genes are not exclusively expressed in LPC, the marked upregulation of a significant number of genes typically expressed in isolated LPS suggest that this cell population may be enriched in patients with AH. Results shown in figure 1 are a subset of genes selected from the microarray previously described (29). Full microarray data are deposited in NCBI's Gene Expression Omnibus (GEO; accession number GSE28619).

Detailed analysis of selected liver progenitor cell markers in patients with alcoholic hepatitis and other liver diseases.

KRT7, PROM1 and *EpCAM* were selected for further analysis as representative genes described in ductular reaction. Expression of these genes was evaluated by real time PCR in the whole series of patients with AH (n=59), in comparison with the expression in tissue from patients with hepatitis C virus (HCV)-induced cirrhosis (HCV-CH) (n=16), chronic hepatitis C (HCV) (n=14) and fragments of normal tissue (n=12). Gene expression of all three genes was significantly higher in patients with AH compared to HCV-induced liver disease, and normal livers (Figure 2a-c). Moreover, although *KRT7, PROM1* and *EpCAM* are expressed in different cell populations in the ductular reaction, there was a significant correlation among the expression of these markers (data not shown). We next investigated if *KRT7, PROM1* and *EpCAM* were differentially expressed in patients with AH according to their survival rate. *KRT7* and *PROM1* expression was higher in patients that died within 90 days after admission (n=16) compared to those that survived (n=43) (Figure 2d-f). There was no difference in *EpCAM* expression between patients with different survival rate (Figure 2f). To further assess if the expression of LPC markers was associated with disease severity, we evaluated the correlation of gene expression of LPC markers with clinical prognostic scores. As shown in figure 2, we found a positive correlation of *KRT7*, but not *PROM-1* and *EpCAM* gene expression with MELD and Maddrey's discriminant function (Figure 2g, h), suggesting that liver damage may promote a ductular reaction that not necessarily correlates with the expansion of more immature progenitor cell markers. These results demonstrate that AH is characterized by an important ductular reaction and that the liver damage may be an important trigger of this event.

Factors Influencing Patient Survival: Role of Progenitor Cell Markers

KRT7, PROM1 and *EpCAM* were found to be upregulated in patients with poor survival rate. Thus, we evaluated if these genes are good prognostic markers of mortality in patients with AH. The univariate analysis identified serum bilirubin levels, INR, serum creatinine levels, ABIC score and MELD at admission and *KRT7 and Prominin-1* expression in liver tissue associated

with 90 day mortality. However, age, *EpCAM* and fibrosis quantification were not associated with mortality. In the multivariate regression analysis, the ABIC score, MELD, *KRT7* and *PROM-1* were the best independent predictors of 90 days mortality (Table 2).

In order to assess if LPC markers are surrogate markers for fibrosis and inflammation, we evaluated liver fibrosis and infiltrating polimorphonuclear cells and we assessed its association with LPC markers and mortality. As shown in supplementary figure 2, fibrosis stage had a low but positive correlation with *KRT7* and *EpCAM* expression. Moreover, morphometric quantification of fibrosis had also a positive correlation with expression of *KRT7* (r=0.28 p=0.048) and *PROM-1* (r=0.3 p=0.031) but not *EpCAM* (r=0.21 p=0.119). However, fibrosis was not found to be an independent predictor of mortality with an area under the ROC curve of 0.59 (0.41-0.75 95% CI) and was unable to stratify patients based on the outcome. The degree of inflammation did not positively correlate with LPC markers (Supplementary Figure 2) nor with mortality.

In order to assess the interactions among LPC markers and their direct association with short-term survival of patients with AH, we fitted a CART model. The best decision tree (Gini´s index 0.35) was constructed with the interaction of two LPC markers, *KRT7* and *EpCAM* (Figure 3). The generated tree identified three sub-populations with different short-term prognosis (expressed as $2^{(-\Delta\Delta CT)}$): a high-risk mortality group characterized with high *KRT7* expression (KRT7≥12.9), with a 90-day mortality of 89%; a group with an intermediate mortality risk (*KRT7* between 12.9 and 3.5 and *EpCAM* ≤2.95), with a 90-day mortality of 33%; and a low-risk mortality group (*KRT7* <12.9 and *EpCAM* >2.95) and (*EpCAM* ≤2.95 and *KRT7* ≤3.5) with a 90-day mortality of 0% and 10% respectively. Importantly, this CART model showed good usefulness estimated by an AUROC of 0.73, CI 95%: 0.60-0.87.

Finally, we investigated if LPC markers provide additional discrimination potential within patients classified by the ABIC score as with intermediate mortality risk (ABIC score 6.71-8.99

or ABIC B) which show a mortality rate of 30%. The area under the ROC for 90-day mortality using *KRT7* was 0.76 (CI 95% 0.60-0.88) (Supplementary figure 3). We chose a cut off value of 6.4 with a sensitivity of 60% and a specificity of 83%. Interestingly, using this cut off value, we found that *KRT7* expression accurately discriminated patients with different short-term survival rate among the group of patients with an ABIC B (86% *vs* 50% survival in patients with KRT7< 6.4 and KRT7≥6.4, respectively; *p=0.026*, Figure 4).

Protein expression of progenitor markers: correlation between immunohistochemistry and gene expression

CART analysis identified *KRT7* and *EpCAM* expression as the most relevant variables to subdivide the population of patients with AH into sub-populations with different outcomes. To assess if real time PCR gene expression correlated with protein expression and ductular reaction cells in AH samples, we performed an immunohistochemistrty analysis of KRT7 and EpCAM. As described previously, we identified three main different cell populations positive for KRT7 or EpCAM according to their morphology and localization in the liver. To reflect this heterogeneity we have developed a score (see Methods section) to quantify the contribution of these different populations within the ductular reaction. As shown in Figure 5, the KRT7 score positively correlated with *KRT7* gene expression (r=0.63; p=0.003) (Figure 5g). *EpCAM* score also showed a positive correlation with EpCAM gene expression (r=0.88; p<0.001) (Figure 5h).

To evaluate if LPC markers reflect different progenitor cell populations or different differentiation states, we performed a dual staining for LPC markers together with a hepatocyte marker. As it is shown in figure 6 there is almost a complete overlay of EpCAM and KRT7 staining in cells of the ductular reaction, suggesting that both markers may be identifying the same cell population. Dual staining for EpCAM and hepatocyte marker has shown that cells with a membranous staining of EpCAM co-express hepatocyte specific mitochondrial antigen,

suggesting that differentiation of EpCAM positive cells occurs during an episode of alcoholic hepatitis. Likewise, KRT7 positive cells showing an immature hepatocyte morphology co-expressed hepatocyte markers.

DISCUSSION

The present study describes the expression of progenitor cell markers in patients with AH and identifies them as prognostic markers predicting short-term mortality. Gene expression analysis showed an important upregulation of progenitor cell markers in patients with AH. A further analysis of three well-known LPC markers, *KRT7, PROM1* and *EpCAM* in a cohort of patients with AH allowed us the possibility of stratifying patients into three groups according to their survival rate. Moreover, the combination of KRT7 expression with the ABIC score discriminate patients with intermediate ABIC values and poor defined survival rate and stratify them according to mortality.

LPC proliferate during the course of chronic liver disease. This proliferation is particularly important in alcoholic liver disease, probably because AH causes profound hepatocellular damage and impairment of hepatocyte proliferation but also because alcohol triggers progenitor cell expansion (24). However, little is known about the extent of progenitor cell proliferation and their role in acute-on-chronic liver failure. Our results are in accordance with previous reports describing an important ductular reaction in patients with alcoholic liver disease (23, 24). Here, we show that a broad number of LPC markers are overexpressed in patients with AH, as assessed by microarray analysis. Moreover, we demonstrate that progenitor cell markers are overexpressed in patients with AH compared to hepatitis C virus induced cirrhosis, suggesting that acute-on-chronic injury may favor progenitor cell expansion. However, the group of patients with HCV-induced cirrhosis had a better liver function than the AH group, so it is plausible that LPC expansion is not a unique feature of AH. Further studies are required to determine if LPC proliferation is increased in AH compared to other etiologies of chronic liver disease.

The main innovative approach of this study is the use of CART as an alternative regression analysis method to evaluate progenitor cells markers interactions as indicators of poor

prognosis in patients with AH. The rationale behind the application of CART in this study is two fold. First, traditional logistic regression methods can be very useful in ranking the variables with statistical significance, but they contain modest information about the impact of the identified variables (30). The elucidation of risk subgroups seems unwieldy with a traditional linear regression analysis and even if estimates for incidences are calculated for all combinations of parameters, the results still refer only to estimates. Conversely, the results of CART analysis come naturally as risk groups that are based on observed incidences and impact of the identified parameters, providing a graphical representation in the form of a decision tree. Second, in contrast to logistic regression, CART considers not only the overall sample of patients, but also, in subsequent steps, relevant subgroups and is therefore better positioned to probe for interactions. Thus, the main benefit of CART is that it can accurately stratify prognostic subgroups based on simple combinations of variables given to semi-automatic driven software.

Several scoring systems have been developed for the prognosis assessment of patients with AH (30-33). Those scores are based on non-invasive parameters and are reliable tools to estimate the severity and prognosis of these patients. However, it is unknown which mechanisms contribute to worsen the prognosis of patients with AH. Gene and protein expression of LPC markers are of limited applicability as a prognostic score tools in the clinical practice, thus, our aim was not to develop a clinically relevant prognostic score but to identify relevant pathophysiological mechanisms influencing the outcome in AH. We propose a simple CART model (Figure 3) based on the combination of two LPC markers that evidences the importance of LPC expansion in AH and allows an accurate discrimination of three prognostic groups of patients with different mortality. First, a low-risk group with an excellent overall 90-day survival of 94%, for whom aggressive therapeutic measures may not be needed. The moderate expression of LPC markers observed in these patients probably reflects a low inflammatory damage and moderate hepatocyte loss combined with an active regeneration

process. Second, a high-risk group of patients with 89% mortality at 90 days. These patients with the highest KRT7 expression and poor prognosis have a prominent inflammatory and necrotic damage with a severe impairment of the hepatocyte regeneration capacity. And third, patients with intermediate prognosis (33% mortality at 90 days) for which aggressive therapeutical approaches are required.

It is especially relevant that *KRT7* expression accurately predicts the 90-day mortality of patients with intermediate mortality as assessed by a moderate ABIC score. This finding may have potential clinical implications since the mortality of moderate ABIC is similar to that observed in the general AH patient population. By using *KRT7* expression we were able to stratify patients with intermediate ABIC score in those having a low mortality rate (14%) and high mortality (50%). Future studies should assess KRT7 expression in a higher number of patients with moderate ABIC score to validate these results. The possibility to identify those patients at higher risk may be useful for optimizing patient management and therapeutic decision-making. Moreover, these observations suggest that liver regeneration and LPC expansion may be an important factor influencing short-term prognosis in AH. Ductular reaction is frequently associated with fibrosis and inflammation in chronic liver disease and may be key factors promoting LPC expansion in AH. Although we show that fibrosis correlates with LPC markers, fibrosis and inflammation are not independent predictors of bad prognosis, and for that reason are not adequate parameters to stratify AH patients according to mortality. It has been previously shown that fibrosis and inflammation trigger LPC expansion, but our results suggest that LPC markers would not be surrogate markers for fibrosis since KRT7, PROM-1 and EpCAM are independent prognostic factors of mortality in our cohort of patients.

Gene expression and immunochemistry results clearly demonstrate that there is an important expansion of KRT7 positive hepatobiliary cells in AH. It is assumed that LPC proliferate to

overcome the impaired regenerative potential of a chronically damaged liver in order to restore the normal liver function. Although our results clearly demonstrate an important LPC expansion, it does not lead to improved liver function, since patients with an increased number of immature hepatobiliary cells show the highest mortality rate. This observation raises the question whether the presence of KRT7 cells correlates with more extensively damaged liver, therefore showing a higher mortality, or whether LPC proliferation has a detrimental impact on liver function. AH is a condition with a florid pro-inflammatory milieu, with expression of chemokines from CXC and CCL families, growth factors and other mediators with unknown effects on LPC proliferation, differentiation or three-dimensional organization. The possibility that in AH there is an induction of LPC proliferation but a deficient differentiation to mature hepatocytes deserves further investigation. In chronic liver disease EpCAM positive hepatobiliary cells derive from LPC that give rise to new hepatocytes generated independently from the exhausted mature hepatocyte pool (19). The presence of a number of EpCAM positive hepatobiliary cells in patients with AH suggests that newly generated hepatocytes are being created. Results obtained with the dual staining of LPC markers suggest that differentiation of EpCAM and KRT7 positive cells is not completely abrogated in AH. However, the major population of EpCAM positive cells in samples from AH patients are small immature cells, and only a small fraction of cells are intermediate hepatobiliary cells with an immature hepatocyte morphology. Unfortunately it is not known the contribution of EpCAM or KRT7 positive cells to the reconstitution of the liver parenchyma in liver regeneration and disease, so from our results it is not possible to envision if the contribution of LPC may be sufficient to maintain liver function in AH patients. It is important to notice that as observed in the CART analysis, high *EpCAM* expression in patients with intermediate *KRT7* expression identify a group of patients with low mortality rate, suggesting a beneficial effect of EpCAM-positive cells on liver function. If EpCAM positive cells sufficiently

contribute to liver regeneration to maintain the liver function in AH deserves further investigation.

In summary, we provide evidence that ductular reaction is a key event in AH and that LPC expansion parallels disease severity in this acute-on-chronic condition. The exploitation of a decision tree model as an alternative approach to classical regression models identified the most relevant LPC markers as prognostic markers in patients with AH. Our results suggest that LPC markers may provide additive value to clinical prognostic scores to discriminate patients with different mortality rate. This study shows a significant cellular response in AH and provides evidences for their utilization in a histological score. Moreover, the correlation of LPC markers with mortality suggest that understanding the mechanisms governing LPC proliferation and differentiation in liver disease may facilitate the design of new therapeutic approaches aiming at promoting liver regeneration in AH.

References

1. Elphick DA, Dube AK, McFarlane E, Jones J, Gleeson D. Spectrum of liver histology in presumed decompensated alcoholic liver disease. Am J Gastroenterol. 2007;102(4):780-8.

2. Teli MR, Day CP, Burt AD, Bennett MK, James OF. Determinants of progression to cirrhosis or fibrosis in pure alcoholic fatty liver. Lancet. 1995;346(8981):987-90.

3. Naveau S, Giraud V, Borotto E, Aubert A, Capron F, Chaput JC. Excess weight risk factor for alcoholic liver disease. Hepatology. 1997;25(1):108-11.

4. Lucey MR, Mathurin P, Morgan TR. Alcoholic hepatitis. N Engl J Med. 2009;360(26):2758-69.

5. Dominguez M, Rincon D, Abraldes JG, Miquel R, Colmenero J, Bellot P, et al. A new scoring system for prognostic stratification of patients with alcoholic hepatitis. Am J Gastroenterol. 2008;103(11):2747-56.

6. Duncan AW, Dorrell C, Grompe M. Stem cells and liver regeneration. Gastroenterology. 2009;137(2):466-81.

7. Zaret KS, Grompe M. Generation and regeneration of cells of the liver and pancreas. Science. 2008;322(5907):1490-4. PMCID: 2641009.

8. Roskams T, Katoonizadeh A, Komuta M. Hepatic progenitor cells: an update. Clin Liver Dis. 2010;14(4):705-18.

9. Eleazar JA, Memeo L, Jhang JS, Mansukhani MM, Chin S, Park SM, et al. Progenitor cell expansion: an important source of hepatocyte regeneration in chronic hepatitis. J Hepatol. 2004;41(6):983-91.

10. Roskams TA, Theise ND, Balabaud C, Bhagat G, Bhathal PS, Bioulac-Sage P, et al. Nomenclature of the finer branches of the biliary tree: canals, ductules, and ductular reactions in human livers. Hepatology. 2004;39(6):1739-45.

11. Turner R, Lozoya O, Wang Y, Cardinale V, Gaudio E, Alpini G, et al. Human hepatic stem cell and maturational liver lineage biology. Hepatology. 2011;53(3):1035-45. PMCID: 3066046.

12. Zhang L, Theise N, Chua M, Reid LM. The stem cell niche of human livers: symmetry between development and regeneration. Hepatology. 2008;48(5):1598-607.

13. Bateman AC, Hubscher SG. Cytokeratin expression as an aid to diagnosis in medical liver biopsies. Histopathology. 2010;56(4):415-25.

14. Libbrecht L, Desmet V, Van Damme B, Roskams T. Deep intralobular extension of human hepatic 'progenitor cells' correlates with parenchymal inflammation in chronic viral hepatitis: can 'progenitor cells' migrate? J Pathol. 2000;192(3):373-8.

15. Ma S, Chan KW, Hu L, Lee TK, Wo JY, Ng IO, et al. Identification and characterization of tumorigenic liver cancer stem/progenitor cells. Gastroenterology. 2007;132(7):2542-56.

16. Suzuki A, Sekiya S, Onishi M, Oshima N, Kiyonari H, Nakauchi H, et al. Flow cytometric isolation and clonal identification of self-renewing bipotent hepatic progenitor cells in adult mouse liver. Hepatology. 2008;48(6):1964-78.

17. Schmelzer E, Zhang L, Bruce A, Wauthier E, Ludlow J, Yao HL, et al. Human hepatic stem cells from fetal and postnatal donors. J Exp Med. 2007;204(8):1973-87.

18. Yamashita T, Ji J, Budhu A, Forgues M, Yang W, Wang HY, et al. EpCAM-positive hepatocellular carcinoma cells are tumor-initiating cells with stem/progenitor cell features. Gastroenterology. 2009;136(3):1012-24. PMCID: 2828822.

19. Yoon SM, Gerasimidou D, Kuwahara R, Hytiroglou P, Yoo JE, Park YN, et al. Epithelial cell adhesion molecule (EpCAM) marks hepatocytes newly derived from stem/progenitor cells in humans. Hepatology. 2011;53(3):964-73.

20. Duret C, Gerbal-Chaloin S, Ramos J, Fabre JM, Jacquet E, Navarro F, et al. Isolation, characterization, and differentiation to hepatocyte-like cells of nonparenchymal epithelial cells from adult human liver. Stem Cells. 2007;25(7):1779-90.

21. Sancho-Bru P, Najimi M, Caruso M, Pawelyn K, Cantz T, Forbes SJ, et al. Stem and progenitor cells for liver repopulation: can we standardize the process from bench to bedside? Gut. 2008.

22. Roskams TA, Libbrecht L, Desmet VJ. Progenitor cells in diseased human liver. Semin Liver Dis. 2003;23(4):385-96.

23. Lowes KN, Brennan BA, Yeoh GC, Olynyk JK. Oval cell numbers in human chronic liver diseases are directly related to disease severity. Am J Pathol. 1999;154(2):537-41. PMCID: 1849988.

24. Jung Y, Brown KD, Witek RP, Omenetti A, Yang L, Vandongen M, et al. Accumulation of hedgehog-responsive progenitors parallels alcoholic liver disease severity in mice and humans. Gastroenterology. 2008;134(5):1532-43.

25. Roskams T, Yang SQ, Koteish A, Durnez A, DeVos R, Huang X, et al. Oxidative stress and oval cell accumulation in mice and humans with alcoholic and nonalcoholic fatty liver disease. Am J Pathol. 2003;163(4):1301-11. PMCID: 1868311.

26. Ginès P AP, Lenz K, Møller S, Moore K, Moreau R, Merkel C, Ring-Larsen H, Bernardi M, Garcia-Tsao G, Hayers P. EASL clinical practice guidelines on the management of ascites, spontaneous bacterial peritonitis, and hepatorenal syndrome in cirrhosis. J Hepatol. 2010;53(3):397-417.

27. Garcia-Tsao G, Sanyal AJ, Grace ND, Carey W. Prevention and management of gastroesophageal varices and variceal hemorrhage in cirrhosis. Hepatology. 2007;46(3):922-38.

28. Runyon BA. Management of adult patients with ascites due to cirrhosis: an update. Hepatology. 2009;49(6):2087-107.

29. Dominguez M LJ, del Castillo AL, Miquel R, Colmenero J, Moreno M, Garcia-Pagán JC, Arroyo V, Gines P, Caballeria J, Bataller R. . Identification of Fn14, a tnf receptor super-family member, as a therapeutic target in patients with alcoholic hepatitis. Hepatology. 2009;50(S4):600A–705A.

30. Muller R, Mockel M. Logistic regression and CART in the analysis of multimarker studies. Clin Chim Acta. 2008;394(1-2):1-6.

31. Dunn W, Jamil LH, Brown LS, Wiesner RH, Kim WR, Menon KV, et al. MELD accurately predicts mortality in patients with alcoholic hepatitis. Hepatology. 2005;41(2):353-8.

32. Forrest EH, Evans CD, Stewart S, Phillips M, Oo YH, McAvoy NC, et al. Analysis of factors predictive of mortality in alcoholic hepatitis and derivation and validation of the Glasgow alcoholic hepatitis score. Gut. 2005;54(8):1174-9. PMCID: 1774903.

33. Maddrey WC, Boitnott JK, Bedine MS, Weber FL, Jr., Mezey E, White RI, Jr. Corticosteroid therapy of alcoholic hepatitis. Gastroenterology. 1978;75(2):193-9.

Figure legends

Figure 1. Heatmap showing hierarchical clustering using expression profiles of the selected genes. Distances are measured using Pearson correlation. Results are expressed as a matrix view of gene expression data where rows represent genes and columns represent hybridized samples. The intensity of each colour denotes the standardized ratio between each value and the average expression of each gene across all samples. Red pixels correspond to an increased abundance of mRNA in the indicated liver biopsy sample, whereas green pixels indicate decreased mRNA levels.

Figure 2. Gene expression analysis of liver progenitor cell markers in liver diseases. Quantitative gene expression analysis of Prominin-1 (a), Keratin 7 (b) and EpCAM (c) in normal livers, hepatitis C virus-induced hepatitis (HCV), HCV-induced cirrhosis (cirrhosis) and alcoholic hepatitis (AH). Expression of Prominin-1 (d), Keratin 7 (e) and EpCAM (f) in patients with short-term survival <90 days and > 90 days. Gene expression values are shown as $2^{(-\Delta\Delta Ct)}$ value. Correlation of Keratin 7 (KRT7) expression value with Maddrey's discriminant function (G) and MELD (h).

Figure 3. Prognostic model of 90-day mortality in patients with alcohol hepatitis generated by CART. The CART tree stratified three groups of patients with different mortality rate according to progenitor cell markers expression. Gene expression values are shown as $2^{(-\Delta\Delta Ct)}$ value. A high risk group (15% of the patients with 89% mortality) with KRT7≥12.9 value, an intermediate risk group (30% of patients with 33% of mortality) with 3.5≤KRT7≤12.9 and EpCAM≤2.95 value, and a low risk group (54% of patients with 6% of mortality) with KRT7<12.9 and EpCAM>2.95 and a second group with EpCAM<2.95 and KRT7≤3.5.

Figure 4. Three-month survival probability of patients with moderate ABIC score according to hepatic gene expression of *Keratin7*. Kaplan-Meier curve showing the three-month

probability of survival in patients with alcoholic hepatitis and ABIC B score according to hepatic gene expression of *Keratin7 (KRT7).*

Figure 5. Immunohistochemistry analysis of liver progenitor cell markers. Representative pictures of keratin 7 staining (a-c) showing biliary structures (a) (magnification x20), isolated cells within the parenchyma (b) (magnification x40), intermediate hepatobiliary cells (c) (magnification x40). Representative pictures of EpCAM staining (d-f) demonstrating presence of ductular structures (d) (magnification x20), cells at the interphase (e) (magnification x40), hepatobiliary cells (f) (magnification x40). Correlation of keratin 7 expression value ($2^{(-\Delta\Delta Ct)}$) with Keratin 7 score (g). Correlation of EpCAM expression value ($2^{(-\Delta\Delta Ct)}$) with EpCAM score (h).

Figure 6. Dual immunostaining of liver progenitor cell markers. (A) Representative pictures of dual immunofluorescent staining of EpCAM, Hepatocyte marker and merge picture (magnification x40). (B) Representative pictures of dual immunofluorescent staining of EpCAM, KRT7 and merge picture (magnification x40). (C) Representative pictures of dual immunofluorescent staining of KRT7, Hepatocyte marker and merge picture (magnification x40).

Acknowledgments

Table 1. Baseline Demographic, Clinical and Biochemical Characteristics of Patients with Alcoholic Hepatitis.

		Study Cohort (n=59)
Age (years)		51 (46-55)
Male (%)		41 (70)
Alcohol Intake (g/day)		100 (80-120)
Corticoids (%)		24 (41)
Biochemical parameters		
Hematocrit (%)		29 (19-35)
Leukocytes $x10^9$ /L		8.0 (6.3-11.1)
Platelets $x10^9$ /L		65 (32-134)
Serum creatinine (mg/dL)		0.8 (0.6-1.0)
Serum albumin (g/dL)		25 (23-31)
Serum bilirubin (mg/dL)		10 (4.2-18.3)
International normalized ratio		1.6 (1.3-1.9)
AST (U/L)		132 (86-182)
ALT (U/L)		54 (34-75)
GGT (U/L)		317 (116-923)
Clinical Scores		
Maddrey's discriminant function		57 (33-77)
MELD score		29 (16-24)
ABIC score		7.65 (6.95-8.47)
ABIC class (%)	A (<6.71)	12 (20)
	B (6.71-8.99)	40 (68)
	C (≥9)	7 (12)
Cirrhosis (yes/no)		(42/17)
Fibrosis (F1/F2/F3-F4) n=57		(3/12/42)
Inflammation (mild/severe) n=57		(41/16)

AST, aspartate aminotranferase; ALT, alanine aminotransferase; MELD, model for end-stage liver disease; ABIC, age, bilirubin, INR, creatinine score.

Table 2. Factors Influencing mortality in Univariate and Multivariate Logistic Regression in Patients with Alcoholic Hepatitis

Univariate Logistic Regression	**OR**	**CI 95%**	**p**
At admission			
Serum bilirubin	1.07	1.01-1.14	0.02
INR	1.98	1.11-3.53	0.03
Serum creatinine	2.19	0.86-5.58	0.09
Age	1.07	0.98-1.18	0.12
ABIC score	2.85	1.84-4.43	<0.001
MELD	1.18	1.09-1.27	<0.001
Fibrosis quantification	1.03	0,95-1,12	0,43
Gene expression			
KRT7	1.18	1.08-1.28	0.001
Prominin-1	1.17	1.08-1.27	0.001
EpCam	1.09	0.93-1.28	0.26
Multivariate Logistic Regression	**OR**	**CI 95%**	**p**
ABIC score	1.75	1.00-3.06	0.04
KRT7	1.14	1.04-1.24	0.004
ABIC score	1.94	1.20-3.12	0.006
Prominin-1	1.14	1.05-1.24	0.002
ABIC score	2.25	1.36-3.71	0.001
EpCam	1.16	0.98-1.36	0.06
MELD	1.19	1.01-1.41	0.03
KRT7	1.22	1.04-1.42	0.01
MELD	1.23	1.06-1.43	0.006
Prominin-1	1.14	0.99-1.30	0.06
MELD	1.21	1.05-1.39	0.006
EpCam	1.19	0.95-1.49	0.12

INR, international normalized ratio; KRT7, keratin 7; EpCam, epithelial cell adhesion molecule; ABIC, age, bilirubin, INR, creatinine score.

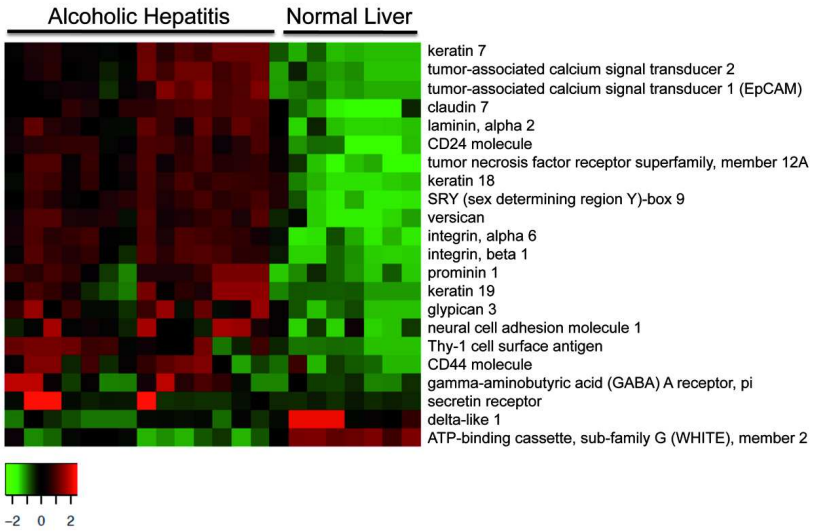


Figure 1

Figure 1. Heatmap showing hierarchical clustering using expression profiles of the selected genes. Distances are measured using Pearson correlation. Results are expressed as a matrix view of gene expression data where rows represent genes and columns represent hybridized samples. The intensity of each colour denotes the standardized ratio between each value and the average expression of each gene across all samples. Red pixels correspond to an increased abundance of mRNA in the indicated liver biopsy sample, whereas green pixels indicate decreased mRNA levels.
124x93mm (300 x 300 DPI)

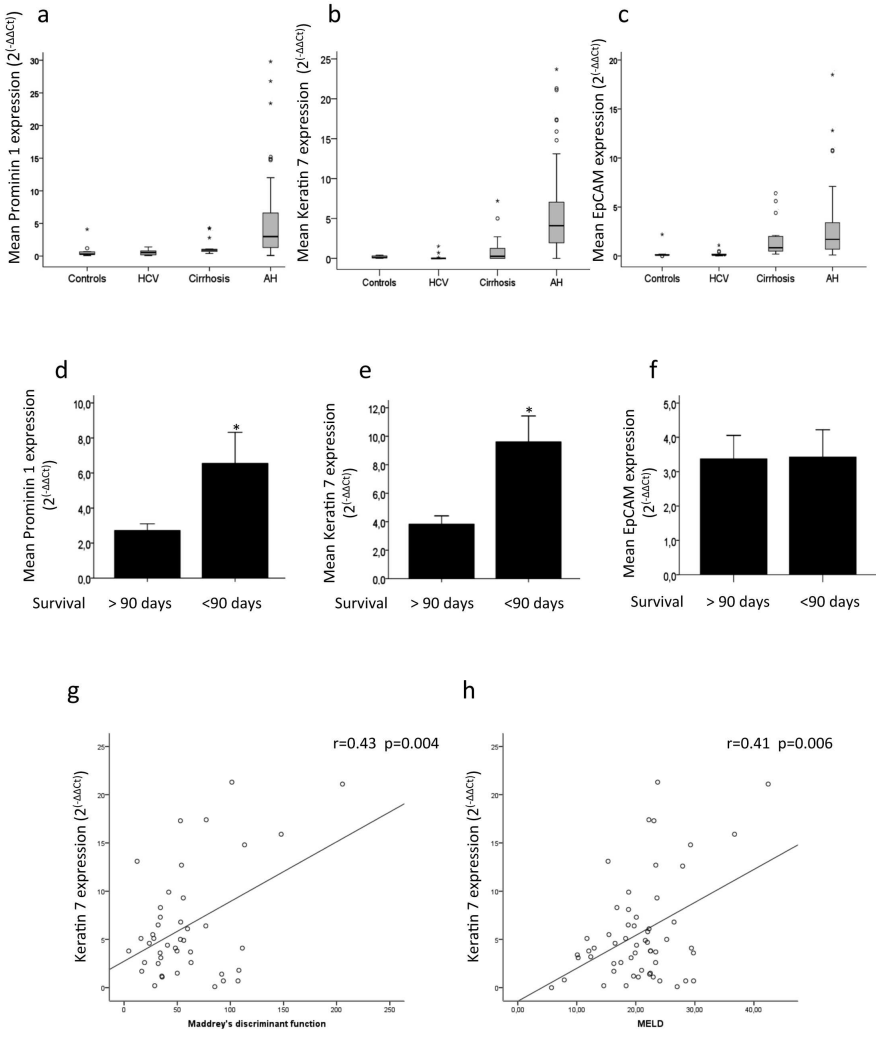


Figure 2

Figure 2. Gene expression analysis of liver progenitor cell markers in liver diseases. Quantitative gene expression analysis of Prominin-1 (a), Keratin 7 (b) and EpCAM (c) in normal livers, hepatitis C virus-induced hepatitis (HCV), HCV-induced cirrhosis (cirrhosis) and alcoholic hepatitis (AH). Expression of Prominin-1 (d), Keratin 7 (e) and EpCAM (f) in patients with short-term survival <90 days and > 90 days. Gene expression values are shown as 2(-ΔΔCt) value. Correlation of Keratin 7 (KRT7) expression value with Maddrey's discriminant function (G) and MELD (h).
190x254mm (300 x 300 DPI)

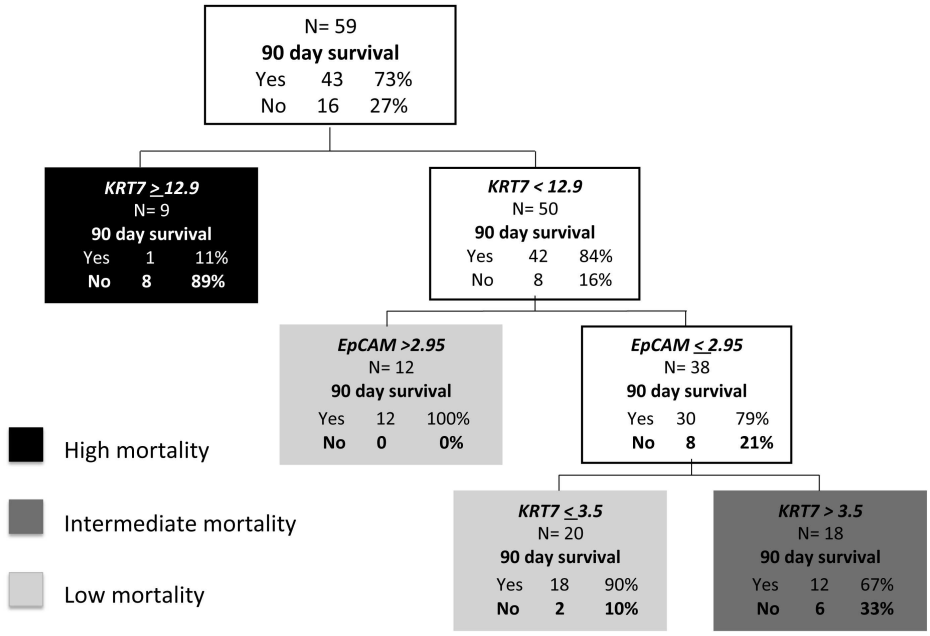
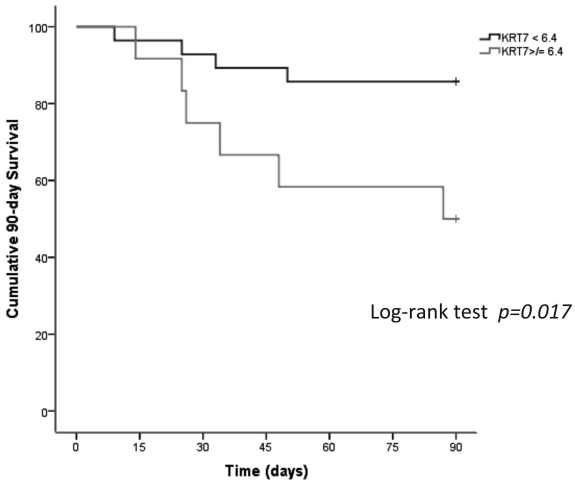


Figure 3

Figure 3. Prognostic model of 90-day mortality in patients with alcohol hepatitis generated by CART. The CART tree stratified three groups of patients with different mortality rate according to progenitor cell markers expression. Gene expression values are shown as 2(-ΔΔCt) value. A high risk group (15% of the patients with 89% mortality) with KRT7≥12.9 value, an intermediate risk group (30% of patients with 33% of mortality) with 3.5≤KRT7≤12.9 and EpCAM≤2.95 value, and a low risk group (54% of patients with 6% of mortality) with KRT7<12.9 and EpCAM>2.95 and a second group with EpCAM<2.95 and KRT7≤3.5.
254x190mm (300 x 300 DPI)



At Risk n=40	day 0	day 15	day 30	day 45	day 60	day 75	day 90
Low KRT7	28	27	26	25	24	24	24
High KRT7	12	11	9	8	7	7	6

Figure 4

Figure 4. Three-month survival probability of patients with moderate ABIC score according to hepatic gene expression of Keratin7. Kaplan-Meier curve showing the three-month probability of survival in patients with alcoholic hepatitis and ABIC B score according to hepatic gene expression of Keratin7 (KRT7).
254x190mm (300 x 300 DPI)

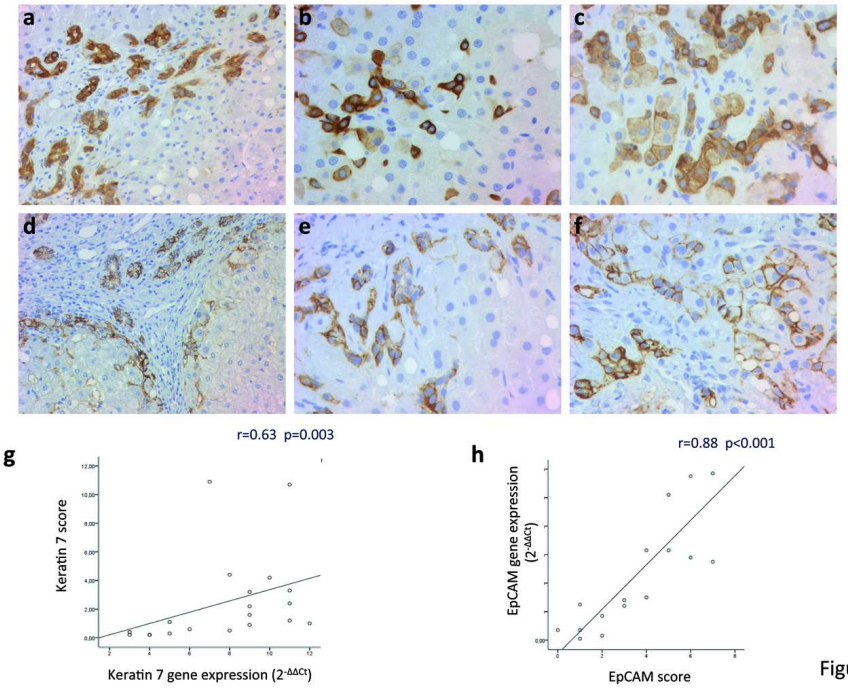


Figure 5

Figure 5. Immunohistochemistry analysis of liver progenitor cell markers. Representative pictures of keratin 7 staining (a-c) showing biliary structures (a) (magnification x20), isolated cells within the parenchyma (b) (magnification x40), intermediate hepatobiliary cells (c) (magnification x40). Representative pictures of EpCAM staining (d-f) demonstrating presence of ductular structures (d) (magnification x20), cells at the interphase (e) (magnification x40), hepatobiliary cells (f) (magnification x40). Correlation of keratin 7 expression value (2(-ΔΔCt)) with Keratin 7 score (g). Correlation of EpCAM expression value (2(-ΔΔCt)) with EpCAM score (h).
124x93mm (300 x 300 DPI)

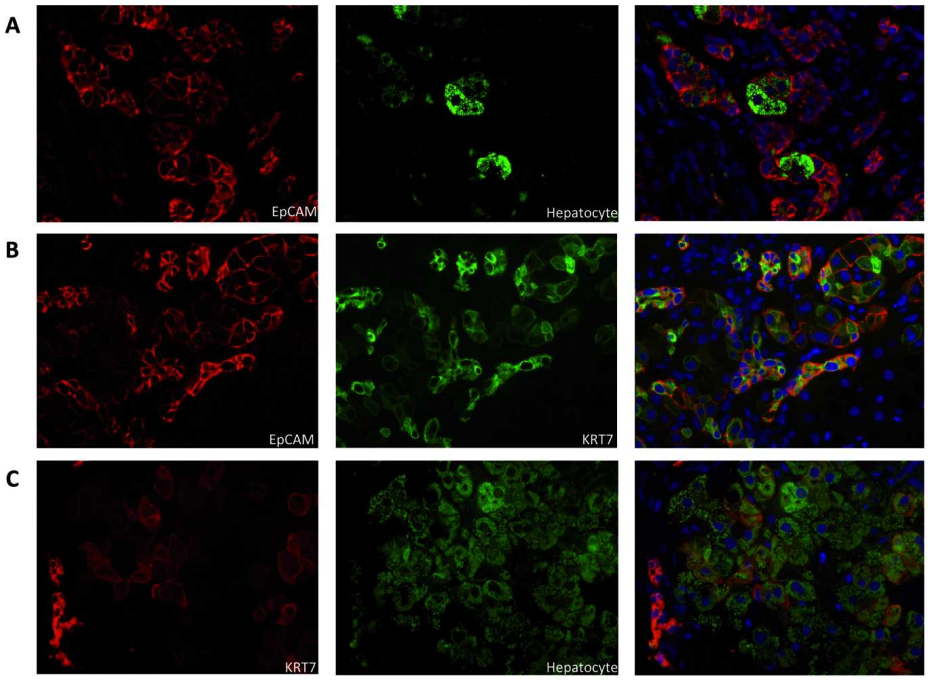


Figure 6

Figure 6. Dual immunostaining of liver progenitor cell markers. (A) Representative pictures of dual immunofluorescent staining of EpCAM, Hepatocyte marker and merge picture (magnification x40). (B) Representative pictures of dual immunofluorescent staining of EpCAM, KRT7 and merge picture (magnification x40). (C) Representative pictures of dual immunofluorescent staining of KRT7, Hepatocyte marker and merge picture (magnification x40).
124x93mm (300 x 300 DPI)

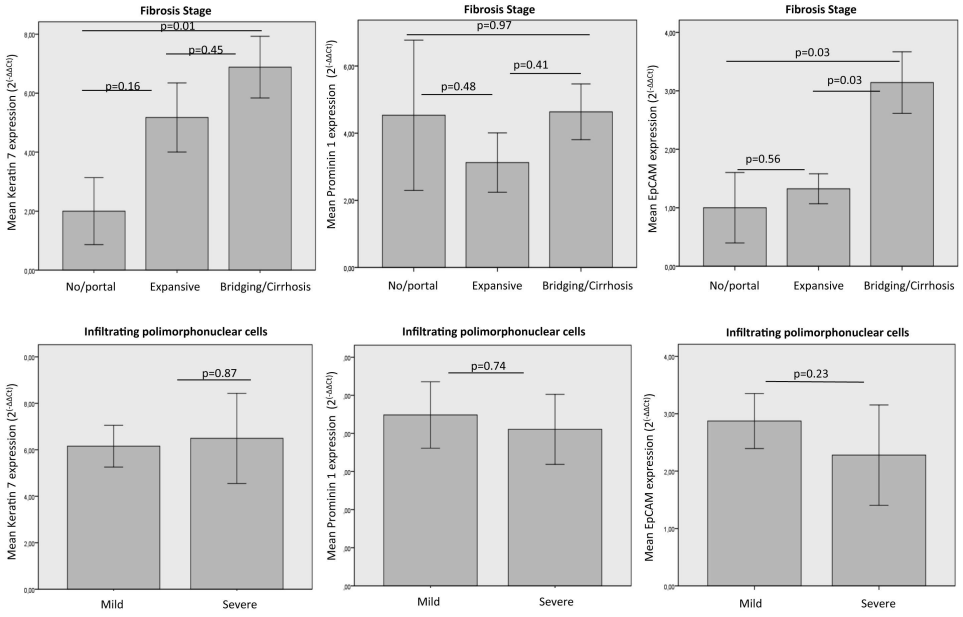
Clinical & Analytical Characteristics of health controls included in the array		
		Controls*
Gender (%)	Male	4 (57)
	Female	3 (43)
Age (years)		51 (43-68)
AST (U/L)		33 (27-63)
ALT (U/L)		35 (22-71)
Bilirubin (mg/dL)		0.6 (0.4-0.8)
γGT (U/L)		47 (13-130)
Albumin (mg/dl)		37 (32-42)
Leukocytes ($x10^9$)		6.9 (4.7-9.7)
Platelets ($x10^9$)		188 (170-219)
INR		1,1 (1.0-1.2)
Creatinine (mg/dl)		0.9 (0.7-1.7)
MELD		5 (2-10)

Numeric data is represented in median (interquartile range [IQR] 25-75).
MELD: Model for End Stage Liver Disease; INR: International Normalized Ratio; AST: Aspartate- Aminotransferase; ALT: Alanin-Aminotrasnferase.
* Liver samples of patient's with liver metastases from colorectal carcinoma except 1 who was affected by primary hyperoxaluria.

Supplementary Figure 1. Clinical and analytical characteristics of healthy controls included in the microarray analysis.

Supplementary Figure 1

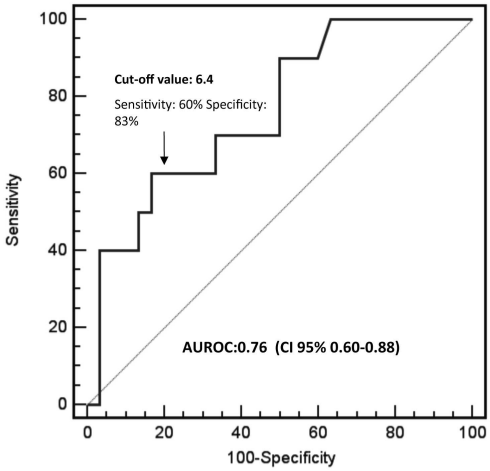
254x190mm (300 x 300 DPI)



Supplementary Figure 2. Correlation of fibrosis stage and degree of inflammation with liver progenitor cell markers. Mean gene expression analysis of Keratin 7 (a), Prominin 1 (b) and EpCAM (c) according to semiquantitative analysis of liver fibrosis stage. Mean gene expression analysis of Keratin 7 (a), Prominin 1 (b) and EpCAM (c) according to histological quantification of infiltrating polimorphonuclear cells.

Supplementary Figure 2

254x190mm (300 x 300 DPI)



Supplementary Figure 3. Receiver-operating characteristic curve (ROC) for 90-day mortality using KRT7 in patients with ABIC B alcoholic hepatitis.

Supplementary Figure 3

254x190mm (300 x 300 DPI)

SUPPLEMENTARY METHODS

Microarray data normalization

Samples were normalized using the guanidine-cytosine content-adjusted robust multiarray algorithm, which computes expression values from probe-intensity values incorporating probe-sequence information (1). Next, we employed a conservative probe-filtering step excluding those probes not reaching a log2 expression value of 5 in at least 1 sample, which resulted in the selection of a total of 19,152 probes out of the original 54,675, set. Differential expression was assessed by using linear models and empirical Bayes moderated t-statistics using LIMMA R-package software (2). Two group comparisons and determinations of false discovery rates (FDR computation using *Benjamini-Hochberg* procedure) were performed and FDR values ≤0.05 were deemed potentially significant and selected for further study. All computations and plots (as heatmaps) were done using R software (2).

Histological analysis and development of KRT7 and PROM-1 score

Dual immunofluorescent staining was performed as serial stainings with EpCAM, KRT7 and the hepatocyte marker clone OCH1e5 (DAKO). Goat anti-mouse Alexa 488 (Invitrogen) and goat anti-mouse CY3 (Jackson Immunoresearch, Suffolk, UK) were used as secondary antibodies. To perform the dual staining, sections were stained with the first primary and secondary antibody. Before performing the second immunostaining, the first complex was stripped by incubating sections for two hours with Restore Western Blot Stripping Buffer (Thermo Scientific, Alcobendas, Spain) followed by overnight incubation with 0.2M Glycine buffer PH:2.5. After stripping, sections were incubated with the second primary and secondary antibodies. As negative controls, all specimens were incubated with primary antibody omission under identical conditions.

Fibrosis quantification

Fibrosis stage was evaluated according to portal fibrosis, expansive fibrosis and bridging fibrosis/cirrhosis (bridging fibrosis and cirrhosis were evaluate together since the size of the sample did not allowed a successful discrimination). The degree of fibrosis was estimated by measuring the percentage of the area stained with tricromic staining. Positive stained area was quantified by a morphometric analysis. Twelve images were obtained with an optic microscope (Nikon Eclipse E600, Nikon Corporation, Japan) at magnification of x20 and analyzed with an image-analysis system (AnalySIS, Soft-Imaging System, Munster, Germany). The positive area was the sum of the area of all positive pixels.

Development of KRT7 and PROM-1 score

Staining for KRT7 was quantified according to 3 parameters punctuated from 0-3 (maximum score = 9): Ductular structures: 0-occasional, 1- proliferation of ductular structures within the portal tract, 2- proliferation of ductular structures within the portal tract and in the interphase between portal tract and parenchyma, 3- ductular reaction intermixed with hepatocyte plates; Isolated cells within the parenchyma: 0- absence, 1- occasional, 2- more than 3 cells per field, 3- apparent presence of cells in the parenchyma; Intermediate hepatobiliary Cells: 0- absence, 1- isolated cells, 2- group of cells, 3- major cell type in some areas of the sample. Staining for EpCAM was quantified according to 3 parameters punctuated from 0-3 (maximum score = 9): Ductular structures: 0- absence, 1- presence of occasional proliferative ductules, 2- apparent proliferation of positive ductular structures, 3- continuous ductular structures occupying most area of the portal tract; Interphase Cells (cells with positive cytoplasm staining at the interphase between the parenchyma and portal tract): 0- absence, 1- occasional, 2- groups of cells lining at the interphase, 3- continuous structures at the interphase intermixed within the parenchyma; Hepatobiliary cells with positive membrane staining: 0- absence, 1- occasional cells, 2- small groups of cells, 3- presence of group of cells covering extensive areas of the parenchyma. The scores where created and evaluated by PSB and RM and externally validated by MC to assess the inter-observer agreement.

Univariate analysis, multivariate analysis and classification and regression tree analysis

In order to assess the potential role of progenitor cells markers in the short-term prognosis of patients with AH, we first performed an exploratory logistic univariate analysis. The p-values for the univariate tests were not corrected for multiple testing, because those tests were considered as exploratory. We also performed a multivariate logistic regression analysis (odds ratio -OR-). The results determined those variables independently associated with the main outcome (after adjusting for the contributions of other variables). Given the small number of death events, to avoid over-fitting, we included only two variables in the multivariate analysis. The SPSS statistical package (SPSS Inc., version 15.0, Chicago, IL) was used for these calculations. To evaluate the interaction of progenitor cells markers over the short-term prognosis of AH patients a classification and regression tree (CART) analysis was performed. We used the CART Pro v6.0 software (Salford systems, San Diego, CA), based on the original Breinman´s code. Briefly, the method allows the construction of inductive decision trees through strictly binary splitting. This algorithm divides each parent node into two homogeneous child nodes by applying yes/no answers at each decision node. The basic tree-generating or "learning" process consists of 1) selecting the most discriminative variable according to an impurity function to partition the data, 2) repeating this partition until child nodes are considered pure enough to be terminal and 3) pruning the resulting tree to avoid overfitting and redundancy. In our study, sensitivity analysis was based on the Gini index, and 10-fold cross validation was performed to assess internal validity. Misclassification costs were set at 1.0. The number of patients in terminal nodes was set to a minimum of 5 and no specific splitters were forced for root or children nodes. Cut-off points for continuous and categorical variables were automatically generated by the model based on statistical cost assumptions. Optimal trees were then selected according to their relative misclassification costs, predictive accuracy and clinical relevance. Third, for prognostic individual testing of every single progenitor cell marker a comparative risk analysis was done using the proposed cut-off

generated by the CART, and a comparison of the Kaplan-Meier method with the log-rank test was performed.

References

1. Wu Z, Irizarri RA, Gentleman R, Murillo FM, Spencer F. A model-based background adjustment for oligonucleotide expression arrays. J Am Stat Assoc. 2004;99:909-17.

2. Smyth GK. Limma: linear models for microarray data. In Bioinformatics and computational biology solutions using R and bioconductor Edited by: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W New York: Springer. 2005:397-420.