Implication of Hepsin from primary tumor in the prognosis of colorectal cancer patients

David Zaragoza-Huesca¹, Andrés Nieto-Olivares², Francisco García-Molina³, Guillermo Ricote¹, Sofía Montenegro¹, Manuel Sánchez-Cánovas¹, Pedro Garrido-Rodríguez¹, Julia Peñas-Martínez¹, Vicente Vicente¹, Francisco Martínez³, María Luisa-Lozano¹, Alberto Carmona-Bayonas¹*, Irene Martínez-Martínez¹, Irene Martínez-Martínez¹*

¹Centro Regional de Hemodonación, Department of Haematology and Medical Oncology, Hospital General Universitario Morales Meseguer, University of Murcia, IMIB-Arrixaca, 30.003, Murcia, Spain.

²Department of Pathology, Hospital General Universitario Morales Meseguer, Murcia, Spain.

³Department of Pathology, Hospital General Universitario Reina Sofía, Murcia, Spain.

⁴Centro de Investigación Biomédica en Red de Enfermedades Raras, U-765-CIBERER, Instituto de Salud Carlos III (ISCIII), 28.029, Madrid, Spain.

*Correspondence to:

Dr. Irene Martínez-Martínez; Address: Centro Regional de Hemodonación, Department of Haematology and Medical Oncology, Hospital General Universitario Morales Meseguer, University of Murcia, IMIB-Arrixaca. Ronda de Garay s/n. 30.003 Murcia, Spain; Tel.: +34-968-341-990; E-mail: immlgi@um.es

Dr. Alberto Carmona-Bayonas; Address: Centro Regional de Hemodonación, Department of Haematology and Medical Oncology, Hospital General Universitario Morales Meseguer, University of Murcia, IMIB-Arrixaca. Ronda de Garay s/n. 30.003 Murcia, Spain; Tel.: +34-968-341-990; E-mail: alberto.carmonabayonas@gmail.com

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT

Background: Hepsin is a type II transmembrane serine protease whose functions include degradation of the extracellular matrix and activation of factor VII. In many tumors, hepsin dysregulation leads to tumor invasion by proteolysis of pericellular components. Colorectal cancer is a tumor with high thrombotic and metastatic risk, but the role of hepsin is unknown.

Objectives: To study the correlations between hepsin expression in the primary tumor and different clinical-histopathological variables in patients with localized and metastatic colorectal cancer.

Methods: A cohort of 287 patients, 169 with localized colorectal cancer and 118 with metastases at diagnosis, was recruited. Tissue microarrays were made from samples at diagnosis of primary tumors and stained with an anti-hepsin antibody. Hepsin levels were determined and the association with clinical-histopathological variables was evaluated using the chi-square and Kruskal-Wallis tests. Time-to-event results and cumulative incidence functions were calculated using Kaplan-Meier and Aalen-Johansen estimators, in combination with Cox and Fine&Gray multivariate models.

Results: In patients with localized colorectal cancer, high intensity hepsin staining was associated with reduced 5-year disease-free survival (p-value = 0.16). In this context, according to multivariable Cox regression, medium and high intensity of hepsin expression versus low intensity was associated with an increased risk of metastatic relapse (hazard ratio 2.83, p-value=0.035, and hazard ratio 3.30, p-value=0.012, respectively), being a more accurate prognostic factor than classic histological variables. Also, in patients with localized colorectal cancer, the cumulative incidence of thrombosis at 5 years increased significantly with higher hepsin levels (p-value = 0.038). According to the multivariate Fine&Gray regression, the medium and high intensity of hepsin with respect to the low intensity, was associated with an increase in thrombotic risk (hazard ratio 7.71, p-value=0.043 and hazard ratio 9.02, p-value=0.028, respectively). This relationship between hepsin and thrombosis was independent of previous tumor relapse (p-value = 0.036). In patients with metastatic colorectal cancer, low intensity of hepsin was associated with a low degree of tumor differentiation (p-value <0.001) and spread to more than one metastatic site (p-value = 0.023).

Conclusion: Hepsin is a potential biomarker of thrombosis and metastasis in patients with localized colorectal cancer. Validation in a larger series and clarification of the underlying mechanism may be helpful in evaluating these patients. In metastatic
patients, Hepsin behaves in a paradoxical way with respect to differentiation and invasion processes.

**Keywords**: Hepsin, colorectal cancer, thrombosis, metastasis, Hepsin paradox.
**Introduction**

Hepsin (HPN) is a type II transmembrane serine protease (TTSP), a family of proteins characterized by a short intracellular amino-terminal domain, a hydrophobic transmembrane domain and an extracellular carboxyl-terminal catalytic domain whose active center is formed by the catalytic triad serine, aspartic acid and histidine (1). HPN is mainly expressed in the liver, although it also exerts its function in other tissues and organs such as adipose tissue, the kidney or the inner ear (2). Among its functions are the regulation of cell growth (3), its participation in the hepatic metabolism of glycogen and lipids (4), the degradation of the extracellular matrix (5) and the activation of procoagulant factors such as factor VII (6,7), which leads us to think of its potential capacity to initiate the coagulation cascade.

In cancer, many extracellular proteases present deregulated levels, which contribute to tumor progression through cleavage of extracellular matrix components (8). Although HPN is part of this proteolytic complex, contributing to degrade the pericellular microenvironment, its net effect on tumorigenesis is complex and context-dependent. Thus, in prostate cancer, HPN has a clear pro-tumorigenic function by promoting invasion processes in certain settings (9). However, other reports in the literature suggest that very high levels of this serine protease may exert paradoxical antitumor effects with limitation of oncogenic signaling and increased autophagy (10). This complex effect explains some discrepant observations reported in the literature. Thus, in ovarian cancer, HPN contributes to tumor progression (11), whereas in endometrial cancer it inhibits tumor cell growth (12). In gastric cancer, elevated HPN levels are associated with a worse prognosis (13), whereas in breast cancer, low HPN levels predict worse survival (14). In addition to these opposing effects, the so-called "HPN paradox" has been described as a dynamic phenomenon whereby, after playing an active role in early tumorigenesis, HPN expression is reduced in more advanced stages to preserve cellular fitness, leading to suppression of its expression in metastatic cells (i.e. limitation of proteolytic excesses once the invasive process is complete) (10).

A limited number of reports in the literature would suggest that HPN might play a more or less crucial role in colorectal cancer (CRC), although the relevance of the enzyme in this context has not been thoroughly evaluated. First, cell invasion processes based on proteolytic systems have been extensively studied, and are relevant in CRC from early transformation to advanced tumorigenesis (15). The individual contribution of HPN in the network of proteases operating in CRC is unknown, although previous work points
to increased serum levels of HPN in patients with advanced versus localized CRC, especially during treatment with chemotherapy (16). Beyond this, the role of HPN along the CRC continuum is unknown, including the association of the protease with relevant clinical events or with thrombotic manifestations (17). Based on the biological actions of the enzyme, we wanted to investigate whether HPN could play a key role in CRC. Specifically, we have evaluated the presence of HPN in CRC samples representative of the evolutionary spectrum of this tumor, to assess the association of its expression with clinicopathological features, survival endpoints, and thrombotic risk.

**Material and Methods**

**Patients**

Patients with localized or metastatic CRC at diagnosis were included in this study. Cases with localized CRC who eventually developed metastases during follow-up were only analyzed within the group of localized neoplasms. Eligibility criteria for patients with localized tumors were: histological diagnosis of colorectal adenocarcinoma, tumors with TNM stage II-III, and microscopic resection with negative margins (R0). Eligibility criteria for metastatic neoplasms were: presence of metastatic disease at diagnosis, and use of at least one line of chemotherapy for advanced disease. Of the overall cohort of 287 patients with CRC specimens at diagnosis, 169 had localized disease and 118 had metastatic disease.

Patients were recruited from two centers (Hospital General Universitario Morales Meseguer [HMM] and Hospital General Universitario Reina Sofia de Murcia), with equivalent diagnostic protocols and with clinical management in a single oncology service, similar reference population, and shared coloproctology committees. Consecutive recruitment was requested in both hospitals.

The project was submitted to and approved by the Clinical Research Ethics Committee of the HMM. The study was conducted in accordance with the Biomedical Research Law 14/2007. Data confidentiality was guaranteed at all times, in accordance with Organic Law 15/1999 on personal data protection, including the rights of access, rectification, cancellation and opposition of data. Informed consent was requested from all patients for their inclusion in the study.

**Tissue microarrays and immunohistochemistry**
Once the primary tumor samples were obtained, on which the diagnosis was made, two pathologists selected the areas with the highest density of tumor cells, and 2 mm diameter cylinders were extracted. Tissue microarrays (TMAs) were constructed from these cylinders with UNITMA equipment (Quick-Ray, Manual Tissue Microarrayer [UNITMA, Co., Ltd, Korea]). Immunohistochemistry of TMAs was performed with a specific anti-HPN antibody (Anti-HPN; HPA006804-100UL; Sigma-Aldrich) using Autostainer Link 48-DAKO automated systems. Each TMA contained patient samples in duplicate; healthy tissue samples from stomach and colon were used as positive controls for HPN staining, and spleen was used as a negative control.

The two pathologists independently evaluated the intensity of HPN immunohistochemical staining semi-quantitatively at three possible intensities: high, medium or low. For reference, controls with positive and negative HPN staining were used, as well as THE HUMAN PROTEIN ATLAS open access database (18). This database has a library of primary tissue biopsies from different cancer types (including CRC) in which HPN immunohistochemical staining is measured according to these three levels of intensity (18).

Clinical and histopathological variables of the patients

The clinical variables evaluated were disease-free survival (DFS), progression-free survival (PFS), overall survival (OS) and cumulative incidence of thrombosis. In patients with localized tumor, DFS was defined as the time interval between localized tumor surgery and metastatic relapse. In patients with advanced tumor, PFS was defined as the interval between initiation of first-line chemotherapy and tumor progression. In both groups of patients, OS and cumulative incidence of thrombosis were defined from the date of disease diagnosis. In all cases, subjects with no events at the last follow-up were censored. Thrombotic disease was diagnosed by imaging techniques (Doppler ultrasound or computed tomography), according to clinical practice.

To model the association between time-to-event endpoints and HPN staining, multivariate models were applied. Model building was performed by understanding the causal mechanisms or sources of bias in CRC. Thus, the covariates for the multivariate model were chosen on theoretical grounds, following literature review (19-24), and taking into account expert opinion. In metastatic tumors, the multivariable model for PFS and OS included HPN staining (specified as a three-level categorical variable), Eastern Cooperative Oncology Group (ECOG) performance status, histologic grade, presence of more than one metastatic site, and specific metastatic sites (peritoneal,
lung, and liver) at diagnosis. In localized tumors, the multivariable model for DFS included HPN staining, histologic grade, lymphovascular invasion, and TNM stage (II versus III) at diagnosis. In both tumor groups, the multivariable model for the cumulative incidence of thrombosis included the covariates histologic grade and lymphovascular invasion at diagnosis. In addition, for localized tumors, it included TNM stage (II versus III) at diagnosis, and for metastatic tumors, the presence of more than one metastatic focus and specific metastatic foci at diagnosis. Other factors of interest were demographic variables, laboratory data (carcinoembryonic antigen [CEA], lactate dehydrogenase, leukocytes, and blood platelets), genetic alterations (KRAS/NRAS/BRAF mutations and microsatellite instability), and other histopathologic variables at diagnosis: perineural invasion, presence and number of tumor deposits distant from the primary tumor, T stage (size and extent of the primary tumor) and N stage (extent of tumor that has spread to nearby lymph nodes).

**Statistics**

The study had a fixed sample size, limited by the availability of tissue samples present in the hospitals where patients were recruited. Therefore, inference should be interpreted according to the width of the confidence intervals (CI). Statistical analysis between HPN immunohistochemical staining and clinicopathologic features was performed in the overall cohort, and stratified for localized and metastatic tumors. Correlation between discrete variables and HPN was performed by chi-square tests (25). In the case of continuous variables, Kruskal-Wallis tests were used (25). Time-to-event outcomes were evaluated by Kaplan-Meier estimator and Log-rank test for trends, and by Cox proportional hazards multivariate regression (26, 27). For the cumulative incidence of thrombosis, Gray's test (28) and the multivariate Fine&Gray test (29) were also used. For the calculation of such incidence, the Aalen-Johansen estimator (30) was also used, taking into account relapse as a competing event (31), to discern whether or not these thromboses were harbingers of relapse (17). All analyses were performed using R-4.1.2 statistical software (32), including the survival (33) and cmprsk (34) packages.

**Results**

*Patients and clinical-histopathologic data*

The clinicopathologic characteristics of the overall cohort, localized patients and metastatic patients are shown in **Table 1**. In patients with localized disease, the
The majority of tumors were T3 (62.7%, N=106/169), followed by T4 (12.7%, N=21/169). Nodal disease (N>0) was present in 53.3% of cases (N=90/169). In patients with advanced disease, the most frequent metastatic location at diagnosis was the liver (71.2% [N=84/118]), and involvement of at least 2 distant organs occurred in 39% (N=46/118) of patients. The remaining characteristics are in accordance with those expected in CRC cohorts treated in clinical practice (Table 1). The median follow-up of patients with localized disease was 41.2 months (range 3.8-190.8), compared to 24.5 months (range 1.4-194) for those with metastatic disease. In localized tumors, 54/169 (32%) patients had a metastatic relapse during follow-up. The median OS in patients with localized tumors was 93 months (95% CI, 83.2-118.2). At the time of analysis, the median OS of subjects with metastatic tumors was 24.7 months (95% CI, 21.5-33.5), and 102/118 (86.4%) had progressed. In patients with localized or metastatic cancer, the cumulative incidence of thrombosis at 5 years was 18.24% (95% CI, 12.69-25.84) and 39.39% (95% CI, 29.05-51.85), respectively (Table 1).

**Supplementary Tables 1, 2 and 3** show the relationship of tumor HPN expression with discrete, time-to-event and continuous variables, respectively.

**Low HPN staining intensity predisposes to poor CRC differentiation**

Analysis of the relationship between HPN expression and histologic grade of tumor differentiation reflects that globally there is an increase in HPN expression from well differentiated to moderately differentiated tumors, which is particularly evident in the case of advanced neoplasms (Figure 1). In the latter, the high intensity of HPN expression ranged from 27% in well differentiated tumors to 78.1% in moderately differentiated tumors (chi-square=31.4, degrees of freedom [df]=4, p-value<0.001) (Supplementary Table 1). This trend however was reversed in the transition from moderately differentiated to poorly differentiated (Figure 1). Furthermore, cases of tumors with low HPN staining intensity increased from those with moderate degree of differentiation, to poorly differentiated (localized: 18.9% versus 50%, chi-square=9.15, df=4, p-value=0.056; metastatic: 12.5% versus 44.4%, chi-square=31.4, df=4, p-value<0.001) (Supplementary Table 1).

**In patients with localized disease, high intensity HPN staining decreases DFS**

As shown in Figure 2, in these patients, HPN staining intensity is associated with worse 5-year DFS rate: 51% (95% CI, 37-71%), 59% (95% CI, 46-75%) and 73% (95% CI, 56-96%) for tumors with high, medium and low staining, respectively (Supplementary Table 2). In multivariable Cox regression, medium and high HPN expression were related to poor prognosis, with hazard ratios (HR) for recurrence of
2.83 (95% CI, 1.07-7.48; p-value=0.035) and 3.30 (95% CI, 1.29-8.40; p-value=0.012), respectively, versus low expression (Table 2). Cox multivariable model reflected that HPN staining provided additional prognostic information to the classical histopathological factors (histological grade, lymphovascular invasion and TNM stage) (Table 2).

The intensity of HPN staining in patients with localized disease at diagnosis was related to the cumulative incidence of thrombosis.

At the time of analysis, 65/287 (22.6%) thrombosis events had been recorded, of which 30/169 (17.8%) occurred in patients with localized disease and 35/118 (29.7%) in metastatic ones. The locations of thrombosis in each cohort are detailed in Table 3. In patients with localized disease, the 5-year cumulative incidence of thrombosis in tumors with high, medium and low HPN staining was 23% (95% CI, 12-33%), 22% (95% CI, 9-33%) and 0%, respectively (Log-rank for trends p-value=0.038 and Gray’s test p-value=0.009; Supplementary Table 2 and Figure 3). In the multivariable Fine&Gray model in patients with localized disease, medium and high HPN staining increased the cumulative incidence of thrombosis, with HR of 7.705 (95% CI, 1.06-55.92; p-value=0.043) and 9.016 (95% CI, 1.27-63.80; p-value=0.028), respectively, versus low staining (Table 4). As in the case of DFS, HPN expression conferred a better prognostic value for thrombosis risk than the other histopathological variables used in the model (Table 4). Supplementary Table 4 shows the detailed and individualized description of the circumstances under which thrombosis occurred in patients with localized disease, in relation to other possible factors that could modulate thrombotic risk other than HPN itself. In patients with localized disease, excluding those who relapsed before the vascular event, 5-year incidence of thrombotic complication when HPN staining intensity was high, medium and low was 16.3% (95% CI, 7.4-25.1%), 7.8% (95% CI, 1.2-14.4%) and 0% (95% CI, 0-0%), respectively (Figure 4). The p-value of the Aalen-Johansen estimator was significant (p-value=0.036).

Association of HPN with the prognosis of metastatic patients and their cumulative incidence of thrombosis.

In the cohort of metastatic patients at diagnosis, at the time of analysis, there had been 102/118 (86.4%) progression events after first-line chemotherapy and 95/118 (80.5%) deaths. HPN staining in the primary tumor was not substantially associated with PFS or OS (Supplementary Tables 2 and 5) nor with thrombotic risk (Supplementary Tables 2 and 5). In contrast, according to multivariable Cox model, other variables such as poor tumor differentiation were related to worse prognosis, with HR of 2.91 for PFS
(95% CI, 1.62-5.25; p-value<0.001), and 3.59 for OS (95% CI, 1.38-6.33; p-value<0.001) (Supplementary Table 5). Similarly, poorly differentiated histological grade was associated with an increased risk of thrombosis, with a HR of 2.85 (95% CI, 1.11-7.33; p-value=0.029) (Supplementary Table 5).

Since the RAS oncogene has been described to promote tumor growth and invasion through HPN activation (35), we wanted to determine whether among metastatic patients with mutated RAS there was correlation between HPN and the time-to-event variables mentioned above. The results reflected that among metastatic patients with RAS gene mutations, HPN staining in the primary tumor was not substantially associated with PFS, OS or cumulative incidence of thrombosis, unlike histopathologic factors that did show that correlation, such as histologic grade, some metastatic locations or ECOG scale (Supplementary Table 6).

Association of HPN with the pattern of tumor dissemination in metastatic patients

Figure 5 shows, for each type of metastatic involvement at diagnosis, the percentage of patients in the different HPN staining intensity groups. HPN staining was not substantially different between patients with or without liver metastases (Figure 5A and Supplementary Table 1). For peritoneal metastases, cases with low HPN staining intensity were higher among patients with such involvement (34.4%) than without it (18.6%), although the differences were not significant (p-value=0.11) (Figure 5C and Supplementary Table 1). Among cases with lung metastasis, high HPN staining decreased its percentage with respect to no lung involvement (32.3% versus 47.1%), in the opposite direction to low staining (32.3% versus 19.5%), although these differences were also not significant (p-value=0.245) (Figure 5B and Supplementary Table 1). Significantly, the percentage of patients with high intensity of HPN staining was lower in those with >1 metastatic foci at diagnosis compared to those with a single metastatic focus (30.4 versus 51.4%, respectively; p-value=0.023). The distribution of patients with low intensity HPN staining was reversed (34.8 versus 15.3%; p-value=0.023) in patients with >1 versus 1 metastasis (Figure 5D and Supplementary Table 1).

Discussion

Proteolytic enzymes are involved in the destruction of the extracellular matrix of CRC, forming a sophisticated network that participates in processes of invasion and metastasis. Therefore, these enzymes could be postulated both as potential biomarkers and as possible therapeutic targets (15). In this study, we have evaluated
the influence of the expression of HPN, a transmembrane serine protease, on the prognosis of localized and metastatic CRC. Our results support that a high expression of HPN in the tumor triples the risk of recurrence after surgery for localized CRC, being a better prognostic factor than other classic histopathological factors (36). Furthermore, in patients with localized disease, increased expression of HPN compared to patients with low staining, raises thrombotic risk, independent of tumor recurrence. In contrast, our results do not support that HPN staining in primary tumor samples influences survival or thrombotic risk in patients with metastatic CRC. In the latter, the paradox of HPN (10) is reflected in the association of lower expression of this serine protease with a lower degree of histological tumor differentiation and greater metastatic spread to distant organs.

CRC is one of the most common tumors and has a variable prognosis depending on the stage, with OS at 5 years being approximately 90% for patients with stage I and slightly more than 10% for patients with stage IV (37, 38). These data reflect that early diagnosis of CRC and the prediction and prevention of metastasis can be decisive for patient survival. Furthermore, CRC, like other tumors of the gastrointestinal tract, is associated with a high risk of thrombotic events, such as pulmonary thromboembolism (39). Thrombosis correlates with increased morbidity and mortality. In fact, some thrombotic markers are associated with a worse prognosis in patients with CRC, such as platelet count (40), D-dimer levels (41), fibrinogen (42), and von Willebrand factor (43).

Therefore, the identification of prognostic markers of metastasis or thrombosis in CRC could help anticipate the appearance of these complications, thus improving the survival of patients. In this context, we highlight that in the cohort of patients with localized disease, the progressive increase in the intensity of HPN staining was accompanied by a lower DFS. This is in agreement with other studies on HPN in prostate and breast cancer, and the underlying functional mechanism could be based on the ability of HPN to promote invasion to distant organs through the disorganization of the basement membrane that surrounds the primary tumor (44-47). Significantly, according to Cox's multivariate model for DFS, HPN was an indicator of relapse risk, and its prognostic value was much more accurate than other histopathologic covariates such as histologic grade, lymphovascular invasion, and TNM stage. Thus, given that there are articles that discredit the usefulness of histopathological variables to predict events related to DFS in CRC (48), we propose HPN as a potential biomarker to predict metastatic relapse.
Another interesting finding among our results in the cohort of patients with localized disease is that the increase in HPN staining in the primary tumor significantly increased the cumulative incidence of thrombosis, with HPN again presenting a better prognostic value than the histopathological covariates. In addition, this relationship was maintained when metastatic recurrence was excluded as a possible underlying cause of thrombosis. The interaction between HPN and coagulation in this type of cancer could be explained by the ability of HPN to activate proteins such as factor VII, XII and IX of the coagulation cascade (6,7,49), although none of these studies have been done in humans. Another key fact to explain the association between HPN and thrombosis is that, although this protein is described as a transmembrane protease, there are studies that have identified it in the serum of patients with CRC (16). Therefore, from the primary tumor, the cells could release the extracellular fraction of HPN into the microvasculature that irrigates the tumor, accessing the bloodstream, where its interaction with coagulation factors would be facilitated.

The effect of HPN on relapse and thrombosis is particularly interesting because, despite tumor resection (50), high levels of HPN at diagnosis maintain an effect on the occurrence of thrombosis and metastatic relapse during follow-up. We do not know the mechanism by which HPN maintains this effect when the tumor is removed, and future studies will be necessary to understand this association. However, a possible explanation could be based on the ability of HPN to degrade the surrounding extracellular matrix (5) and promote tumor cell motility and invasion (51) in areas of healthy cells, which could make total tumor resection difficult. Another possible explanation for the prothrombotic effect is based on the previous arguments about the ability of HPN to enter the circulation (16), so there would be a postoperative soluble fraction that could promote a state of hypercoagulability.

In patients with metastatic CRC at diagnosis, the relationship of HPN to tumor histologic grade and metastatic spread supported "the HPN paradox", according to which tumors have developed a precise spatiotemporal restriction of HPN overexpression (10). Therefore, our results showed that high expression of HPN contributed to achieving moderate tumor differentiation. However, HPN expression decreased in poorly differentiated tumors, with patients with lower HPN staining being the most abundant in this histological grade group. In addition, we found that in those subjects with two or more organs affected by metastases, the expression of HPN was significantly reduced with respect to patients with less metastatic dissemination. In summary, our data argue that as CRC becomes more differentiated from healthy tissue and expands its range of invaded distant organs, HPN expression levels decrease,
possibly because this serine protease is no longer required. An example of utility-dependent tumor expression of HPN occurs in prostate cancer, where this "HPN paradox" was described (52). In the latter, while primary tumors increased HPN levels to promote tumor progression, cells that reached distant tissues and gave rise to metastases reduced their HPN levels. In this case, HPN would no longer be necessary in an environment in which the tumor cell seeks to adhere to the new surrounding matrix to form a new tumor niche, instead of degrading it (52).

When comparing patients with metastatic and localized disease, we found no differences in the levels of HPN staining (Supplementary Table 1), but it should be noted that only in localized neoplasms did we find prognostic value for HPN. We do not know why the predictive effect of HPN on thrombosis and metastasis disappears in patients with metastatic disease. We can hypothesize that in primary tumors that have already metastasized, HPN would lose much of its proinvasive utility and, consequently, its relationship with tumor progression would also be lost. In addition, for both metastatic relapse and thrombosis, our analysis does not include the potential presence of HPN in plasma (16) or metastatic locations (53). In this sense, in patients with metastatic cancer, by considering only the expression of HPN in the primary tumor, we could be underestimating the total levels of HPN in the time-to-event analysis.

Among the limitations of this study, the first is the unique origin of the recruited patients, since they all belong to the same geographical area (Murcia Region, Spain), so the extrapolation of these results to cohorts from other distant geographical areas must be confirmed. The second limitation refers to the results regarding HPN as a thrombotic biomarker. Although the differences in the cumulative incidence of thrombosis are significant according to the expression of this protein in the primary tumor biopsy, thrombosis has a multifactorial etiology that involves not only genetic factors, but also numerous environmental factors that could interfere with the usefulness of HPN as a biomarker.

In conclusion, our article describes HPN as a prognostic marker of metastatic recurrence and thrombosis in patients with localized CRC. If validated in an independent cohort, the ability of HPN to predict metastatic relapse or thrombosis could contribute to the prevention of these complications, which could be decisive for the survival of patients with localized CRC. In addition, in metastatic cancer, HPN expression in the primary tumor appears to be subjected to paradoxical regulation in the processes of tumor differentiation and invasion of distant organs.
Acknowledgements
The authors wish to thank Alberto Martínez for their technical assistance.

Funding
This work was funded by Instituto de Salud Carlos III (ISCIII) (Grant Number PI17/00050 & FEDER, PI21/00210 & FEDER), and the generous donations to Crowdfunding Precipita (FECYT).

David Zaragoza-Huesca holds a predoctoral FPU grant from the Spanish Ministry of Science and Innovation, Pedro Garrido-Rodríguez holds a predoctoral contract from CIBERER and Julia Peñas-Martínez holds a predoctoral grant from the Spanish Society of Thrombosis and Haemostasis.

Ethics Approval and Consent to Participate
The project was submitted to and approved by the Clinical Research Ethics Committee of the HMM. The study was conducted in accordance with the Biomedical Research Law 14/2007. Data confidentiality was guaranteed at all times, in accordance with Organic Law 15/1999 on personal data protection, including the rights of access, rectification, cancellation and opposition of data. Informed consent was requested from all patients for their inclusion in the study.

Consent for Publication
Not applicable.

Conflict of Interest Statement
The authors declare to have no conflict of interest. The funders had no role in study design, data collection, analysis and interpretation, or manuscript writing and submission.

Data availability
The original data are available upon request to the authors.
Bibliography


10. Willbold R, Wirth K, Martini T, Sültmann H, Bolenz C, Wittig R. Excess hepsin proteolytic activity limits oncogenic signaling and induces ER stress and


33. Therneau TM, Lumley T. Package ‘survival’ 2016. URL: https://cran.r-project.org/package=survival.


43. Wang WS, Lin JK, Lin TC, Chiou TJ, Liu JH, Yen CC, Chen PM. Plasma von Willebrand factor level as a prognostic indicator of patients with metastatic


Figure captions

**Figure 1. Distribution of Hepsin staining intensity in patients with different grades of tumor differentiation.** A) Percentage of patients in the overall cohort with different Hepsin staining intensity in each of the three grades of tumor histologic differentiation. B) Percentage of localized patients with different Hepsin staining intensity in each of the three grades of tumor histologic differentiation. C) Percentage of metastatic patients with different Hepsin staining intensity in each of the three grades of histologic tumor differentiation. *HPN*: Hepsin.

**Figure 2. Disease-free survival of localized patients at diagnosis according to Hepsin staining.** Disease-free survival is shown for the three groups of patients with different hepsin levels at diagnosis. Time on the X-axis corresponds to months of follow-up since diagnosis. The Log-rank test for trends gives information on the degree of significance of the differences between the three groups of patients. At the bottom, the patients who may suffer metastatic relapse after a given time from the start of follow-up are shown. *p*: p-value; *tft*: test for trends; *HPN*: Hepsin; *n (%)*: number of patients (percentage).

**Figure 3. Cumulative incidence of thrombosis of localized patients at diagnosis according to Hepsin staining.** The cumulative incidence of thrombosis is shown for the three groups of patients with different hepsin levels at diagnosis. The time on the X-axis corresponds to the months of follow-up since diagnosis. The log-rank test for trends and Gray's test give information on the degree of significance of the differences between the three groups of patients. At the bottom, the patients who may suffer thrombosis after a given time from the start of follow-up are shown. *p*: p-value; *tft*: test for trends; *HPN*: Hepsin; *n (%)*: number of patients (percentage).

**Figure 4. Cumulative incidence of thrombosis independent of relapse according to Hepsin staining in patients with localized cancer at diagnosis.** Time on the X-axis corresponds to the months of follow-up since tumor diagnosis. The p-value from Aalen-Johansen estimator gives information on the degree of significance of the differences between the three groups of patients. *p*: p-value of the Aalen-Johansen estimator; *HPN*: Hepsin.

**Figure 5. Distribution of hepsin staining intensity in patients with different types of metastatic involvement at diagnosis of colorectal tumor.** A) Liver metastasis. B) Lung metastasis. C) Peritoneal metastasis. D) More than one metastatic location. *HPN*: Hepsin.
# Tables

Table 1. Clinical and histopathologic characteristics of patients from the study.

*N*: Number of patients; *ECOG*: Eastern Cooperative Oncology Group; *CI*: Confidence interval; *LL*: Lower limit; *UL*: Upper limit; *T stage*: Size and extent of primary tumor; *N stage*: Extent of tumor that has spread to nearby lymph nodes.

<table>
<thead>
<tr>
<th>Clinical/histopathological features</th>
<th>Overall cohort N=287</th>
<th>Localized patients N=169 (58.9%)</th>
<th>Metastatic patients N=118 (41.1%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age at diagnosis (Range) (years)</td>
<td>66 (22-88)</td>
<td>64 (22-83)</td>
<td>68 (26-88)</td>
</tr>
<tr>
<td>Males: N (%)</td>
<td>187 (65.2)</td>
<td>109 (64.5)</td>
<td>78 (66.1)</td>
</tr>
<tr>
<td>Median follow-up (Range) (months)</td>
<td>35.1 (1.4-194)</td>
<td>41.2 (3.8-190.8)</td>
<td>24.5 (1.4-194)</td>
</tr>
<tr>
<td>ECOG at diagnosis &lt;2: N (%)</td>
<td>237 (82.6)</td>
<td>152 (89.9)</td>
<td>85 (72)</td>
</tr>
<tr>
<td>Median survival (95% CI, LL-UL) (months)</td>
<td>64.5 (55.4-82.9)</td>
<td>93 (83.2-118.2)</td>
<td>24.7 (21.5-33.5)</td>
</tr>
<tr>
<td>Localized patients at diagnosis who undergo metastatic relapse during follow-up: N (%)</td>
<td>NA</td>
<td>54 (32)</td>
<td>NA</td>
</tr>
<tr>
<td>Patients with disease progression after 1st line chemotherapy: N (%)</td>
<td>145 (50.5)</td>
<td>43 (25.4)</td>
<td>102 (86.4)</td>
</tr>
<tr>
<td>5-year cumulative incidence of thrombosis (%; 95% CI, LL-UL)</td>
<td>25.91; 20.49-32.43</td>
<td>18.24; 12.69-25.84</td>
<td>39.39; 29.05-51.85</td>
</tr>
</tbody>
</table>

**Primary tumor**

| Ascending colon: N (%) | 83 (28.9) | 48 (28.4) | 35 (29.7) |
| Descending colon: N (%) | 14 (4.9) | 9 (5.3) | 5 (4.2) |
| Transverse colon: N (%) | 15 (5.2) | 10 (5.9) | 5 (4.2) |
| Sigmoid colon: N (%) | 67 (23.3) | 36 (21.3) | 31 (26.3) |
| Rectal: N (%) | 99 (34.5) | 64 (37.9) | 35 (29.7) |
| Multiple synchronous locations: N (%) | 9 (3.1) | 2 (1.2) | 7 (5.9) |

**Localization of metastases at diagnosis**

<p>| Liver: N (%) | NA | NA | 84 (71.2) |</p>
<table>
<thead>
<tr>
<th>Lung: N (%)</th>
<th>NA</th>
<th>NA</th>
<th>31 (26.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peritoneum: N (%)</td>
<td>NA</td>
<td>NA</td>
<td>32 (27.1)</td>
</tr>
<tr>
<td>Other affectations: N (%)</td>
<td>NA</td>
<td>NA</td>
<td>4 (3.4)</td>
</tr>
<tr>
<td>More than one metastatic site: N (%)</td>
<td>NA</td>
<td>NA</td>
<td>46 (39)</td>
</tr>
</tbody>
</table>

**Histological grade at diagnosis**

<table>
<thead>
<tr>
<th>Well differentiated: N (%)</th>
<th>127 (44.3)</th>
<th>90 (53.3)</th>
<th>37 (31.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moderately differentiated: N (%)</td>
<td>85 (29.6)</td>
<td>53 (31.4)</td>
<td>32 (27.1)</td>
</tr>
<tr>
<td>Poorly differentiated: N (%)</td>
<td>28 (9.8)</td>
<td>10 (5.9)</td>
<td>18 (15.3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lymphovascular invasion at diagnosis: N (%)</th>
<th>123 (42.9)</th>
<th>76 (45)</th>
<th>47 (39.8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perineural invasion at diagnosis: N (%)</td>
<td>57 (19.9)</td>
<td>33 (19.5)</td>
<td>24 (20.3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T Stage &gt;2 at diagnosis: N (%)</th>
<th>191 (66.6)</th>
<th>127 (75.1)</th>
<th>85 (72)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Stage &gt;0 at diagnosis: N (%)</td>
<td>139 (48.4)</td>
<td>90 (53.3)</td>
<td>49 (41.5)</td>
</tr>
</tbody>
</table>

**HPN staining intensity at diagnosis**

<table>
<thead>
<tr>
<th>High: N (%)</th>
<th>122 (42.5)</th>
<th>71 (42)</th>
<th>51 (43.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium: N (%)</td>
<td>104 (36.2)</td>
<td>64 (37.9)</td>
<td>40 (33.9)</td>
</tr>
<tr>
<td>Low: N (%)</td>
<td>61 (21.3)</td>
<td>34 (20.1)</td>
<td>27 (22.9)</td>
</tr>
</tbody>
</table>

**Table 2. Multivariate Cox regression model for disease-free survival in localized patients at diagnosis.** The risk of relapse is calculated according to Hepsin staining intensity, adjusting this calculation by adding different histopathological covariates to the model. **DFS**: disease-free survival; **Coef**: Cox regression coefficient; **exp (coef)**: hazard ratio; **CI**: confidence interval; **LL**: lower limit; **UL**: upper limit; **se (coef)**: standard error of Cox regression coefficient; **p**: p-value; **HPN**: Hepsin; **TNM stage II**: T3/4, N0; **TNM stage III**: T1/2/3/4, N1/2.

<table>
<thead>
<tr>
<th>Multivariate Cox regression for DFS</th>
<th>coef</th>
<th>exp(coef)</th>
<th>LL 95% CI of exp(coef)</th>
<th>UL 95% CI of exp(coef)</th>
<th>se (coef)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPN Medium</td>
<td>1.04</td>
<td>2.84</td>
<td>1.08</td>
<td>7.49</td>
<td>0.50</td>
<td>0.035</td>
</tr>
<tr>
<td>HPN High</td>
<td>1.20</td>
<td>3.30</td>
<td>1.30</td>
<td>8.41</td>
<td>0.48</td>
<td>0.012</td>
</tr>
</tbody>
</table>
Table 3. Location of thrombosis depending on the tumor stage. All of them were deep venous thrombosis, excepting four of them, which were arterial thrombosis (localized: x2 splanchnic; metastatic: x1 splanchnic, x1 femoral). N (%): number of patients (percentage); PE: pulmonary embolism; DVT: deep venous thrombosis; NOS: not otherwise specified. Note: calf vein includes: anterior tibial/ posterior tibial/ fibular veins.

<table>
<thead>
<tr>
<th>Overall, N (%)</th>
<th>Localized, N (%)</th>
<th>Metastatic, N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head &amp; neck</td>
<td>4 (6.2)</td>
<td>2 (6.6)</td>
</tr>
<tr>
<td>Head &amp; neck + PE</td>
<td>3 (4.6)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Catheter-related</td>
<td>2 (3.1)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Catheter-related + PE</td>
<td>1 (1.5)</td>
<td>1 (3.3)</td>
</tr>
<tr>
<td>Splanchnic</td>
<td>9 (13.8)</td>
<td>6 (20)</td>
</tr>
<tr>
<td>Splanchnic + PE</td>
<td>3 (4.6)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Femoral</td>
<td>10 (15.4)</td>
<td>4 (13.3)</td>
</tr>
<tr>
<td>Femoral + PE</td>
<td>2 (3.1)</td>
<td>1 (3.3)</td>
</tr>
<tr>
<td>Calf vein</td>
<td>2 (3.1)</td>
<td>1 (3.3)</td>
</tr>
<tr>
<td>Calf vein + PE</td>
<td>1 (1.5)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Lower extremity, NOS</td>
<td>5 (7.7)</td>
<td>5 (16.7)</td>
</tr>
<tr>
<td>DVT, NOS</td>
<td>1 (1.5)</td>
<td>1 (3.3)</td>
</tr>
<tr>
<td>PE, NOS</td>
<td>22 (33.8)</td>
<td>9 (30)</td>
</tr>
<tr>
<td>Total</td>
<td>65 (100)</td>
<td>30 (100)</td>
</tr>
</tbody>
</table>

Table 4. Multivariate Fine&Gray regression model for the cumulative incidence of thrombosis in localized patients at diagnosis. The risk of thrombosis is calculated according to Hepsin staining intensity, adjusting this calculation by adding different histopathological covariates to the model. Coef: Cox regression coefficient; exp (coef): hazard ratio; CI: confidence interval; LL: lower limit; UL: upper limit; se (coef): standard error of Cox regression coefficient; p: p-value; HPN: Hepsin; TNM stage II: T3/4, N0; TNM stage III: T1/2/3/4, N1/2.
## Multivariate Fine-Gray regression for cumulative incidence of thrombosis

<table>
<thead>
<tr>
<th>Regressions</th>
<th>coef</th>
<th>exp(coef)</th>
<th>LL 95% CI of exp(coef)</th>
<th>UL 95% CI of exp(coef)</th>
<th>se (coef)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPN Medium</td>
<td>2.04</td>
<td>7.71</td>
<td>1.06</td>
<td>55.92</td>
<td>1.01</td>
<td>0.043</td>
</tr>
<tr>
<td>HPN High</td>
<td>2.20</td>
<td>9.02</td>
<td>1.27</td>
<td>63.80</td>
<td>0.998</td>
<td>0.028</td>
</tr>
<tr>
<td>HPN Low (reference)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Moderately differentiated histological grade</td>
<td>0.10</td>
<td>1.11</td>
<td>0.54</td>
<td>2.29</td>
<td>0.37</td>
<td>0.780</td>
</tr>
<tr>
<td>Poorly differentiated histological grade</td>
<td>0.26</td>
<td>0.77</td>
<td>0.08</td>
<td>7.18</td>
<td>1.14</td>
<td>0.820</td>
</tr>
<tr>
<td>Well differentiated histological grade (reference)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lymphovascular invasion</td>
<td>0.37</td>
<td>1.45</td>
<td>0.68</td>
<td>3.07</td>
<td>0.38</td>
<td>0.330</td>
</tr>
<tr>
<td>Absent lymphovascular invasion (reference)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TNM stage III</td>
<td>-0.56</td>
<td>0.57</td>
<td>0.28</td>
<td>1.17</td>
<td>0.36</td>
<td>0.130</td>
</tr>
<tr>
<td>TNM stage II (reference)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Disease-free survival probability

Log-rank, tft
p = 0.16

Number at risk: n (%)
Gray's test

\[ p = 0.009 \]

Log-rank, tft

\[ p = 0.038 \]
Aalen-Johansen estimator: $p = 0.036$