Published in final edited form as:

J Gastrointest Dig Syst. ; Suppl 12: . doi:10.4172/2161-069X.S12-001.

Phenotypic Categorization and Profiles of Small and Large Hepatocellular Carcinomas

Petr Pancoska¹, Sheng-Nan Lu², and Brian I Carr^{3,*}

¹Center for Craniofacial and Dental Genetics, University of Pittsburgh, Pittsburgh, PA, USA

²Division of Gastroenterology, Department of Internal Medicine, Chang Gung Memorial Hospital, Kaohsiung Medical Center, Chang Gung University, Kaohsiung, Taiwan

³Department of Liver Tumor Biology IRCCS de Bellis, National Institute for Digestive Diseases, Castellana Grotte, BA, Italy

Abstract

We used a database of 4139 Taiwanese HCC patients to take a new approach (Network Phenotyping Strategy) to HCC subset identification. Individual parameters for liver function tests, complete blood count, portal vein thrombosis, AFP levels and clinical demographics of age, gender, hepatitis or alcohol consumption, were considered within the whole context of complete relationships, being networked with all other parameter levels in the entire cohort. We identified 4 multi-parameter patterns for one tumor phenotype of patients and a separate 5 multi-parameter patterns to characterize another tumor phenotype of patterns. The 2 subgroups were quite different in their clinical profiles. The means of the tumor mass distributions in these phenotype subgroups were significantly different, one being associated with larger (L) and the other with smaller (S) tumor masses. These significant differences were seen systematically throughout the tumor mass distributions. Essential and common clinical components of L-phenotype patterns included simultaneously high blood levels of AFP and platelets plus presence of portal vein thrombosis. S included higher levels of liver inflammatory parameters. The 2 different parameter patterns of L and S subgroups suggest different mechanisms; L, possibly involving tumor-driven processes and S more associated with liver inflammatory processes.

Keywords

HCC; Tumor mass; Portal vein thrombosis; AFP

Introduction

The prognosis and choice of treatments in patients who have hepatocellular carcinoma (HCC) has long been recognized to depend both on tumor factors as well as liver factors and was the basis for the first published classification scheme of Okuda [1]. This is because HCCs usually arise in a liver that has been chronically diseased (hepatitis or cirrhosis from hepatitis or other causes, or both) [2–6]. Many more complex classification and prognostication schemes have since been published, all of which take these 2 broad

Copyright: © 2013 Pancoska P, et al.

*Corresponding author: Brian I Carr, Department of Liver Tumor Biology IRCCS de Bellis, National Institute for Digestive Diseases, Castellana Grotte (BA), Italy, brianicarr@hotmail.com.

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

categories of factors into account, and patients can die either from tumor growth or liver failure. However, there are additional layers of complexity that need to be taken into consideration. Thus, quite large HCCs can arise in apparently normal (non-cirrhotic liver). Furthermore, many small HCCs do not seem to grow into large HCCs, whereas others do so. Thus, some small HCCs may stay small and others are precursors of larger HCCs. Since a patient can present at any random part of their HCC disease growth process, it is usually difficult to know at what point in their disease process they have been diagnosed. Given the suspicion that the diagnosis of HCC carries within it several subsets of disease, we recently used a tercile approach, to identify HCC subsets at the extreme wings of an HCC patient cohort that had been ordered according to tumor size and then trichotomized into tumor size terciles [7,8]. We found that on the extreme terciles, there was a relationship between plasma platelet numbers and HCC size. This likely reflected that small HCCs arising in cirrhotic liver for which thrombocytopenia is a surrogate [9] and a larger tumor size tercile without thrombocytopenia. However, it still left the central part of the tumor/disease continuum uncharacterized and unordered into subsets. Furthermore, we also showed a relationship between blood alpha-fetoprotein (AFP) levels, a marker of HCC growth, and blood total bilirubin levels, in a large part of the cohort [10]. This led support, as has evidence of others [11–13], that HCC may not only arise and grow in a cirrhotic milieu, but may even depend on signals from that micro-environment for its biology. Given this clinical HCC heterogeneity, it seems that a single approach doesn't work for individual prognostic factors, probably because of the absence of significant sub-subset patient separation. In addition, some parameters such as AFP can be elevated in either small or large HCCs.

We reasoned that attempts to extract new information cannot rely only on standard clinical data, but rather upon processing relationships between the data (Supplementary). In this report, we have taken a different approach to identify phenotypically different HCC patients groups. We first transformed the raw clinical screening data into a new form, considering in full the individual parameters within the whole context of complete relationships to all other parameter levels. After this transformation, individual parameters were not treated as single entries into the analysis, but were each considered as a parameter within the whole clinical context (liver function tests, presence of cirrhosis or hepatitis, inflammation and different manifestations of tumor growth-size, number of tumor nodules, presence of PVT), with considerations of age and gender.

Methods

Patient clinical data

Clinical practice data, recorded within Taiwanese HCC screening program, was prospectively collected on newly-diagnosed HCC patients and entered into a database that was used for routine patient follow-up. Data included: Baseline CAT-scan characteristics of maximum tumor diameter and number, presence or absence of PVT; Demographics (gender, age, alcohol history, presence of hepatitis B or C); Complete blood counts (hemoglobin, platelets, INR); blood AFP and routine blood liver function tests, (total bilirubin, AST and ALT, albumin) (Tables 1 and 2). HCV patients had HCV serum antibodies. HBV patients had HBV serum antigen. Alcohol was determined as daily consumption >10 years. The retrospective analysis was done under a university IRB-approved analysis of de-identified HCC patients.

Patient profiles

We developed a Network Phenotyping strategy (NPS), a graph-theory based approach [10], allowing personalized processing of complex phenotypes, with explicit consideration of

functional parameter correlations and interdependencies. NPS was applied here to integrate the data of all 4139 HCC patients.

There were no missing data in this data set. Individual patient profiles were created, in which each of 15 parameters was assessed in the context of all the other parameters for that same patient and processed by NPS approach. The technical details of NPS are presented in the Appendix. Here we summarize the concrete steps and their results:

Step 1—To reduce the complexity of the relationships that needs to be considered in the analysis, we considered correlations between blood liver function and hematological parameters. Out of 8 liver function parameters, we found 4 unique pairs that showed the most correlated and significant trends in their values. Some of these 4 were also strongly correlated in our previous work [10]. While the selection of the 4 parameter pairs with the strongest correlations amongst all >20,000 possible was done using just a maximal cut mathematical algorithm [14], these 4 unique pairs were inter-related through established underlying functional processes: total blood bilirubin/prothrombin time (a measure of liver function), SGOT/SGPT (a measure of liver inflammation) and AFP and blood platelet counts (reflections of tumor growth) [7].

Step 2—We continued by transforming the original patient data into a form of “levels”. This step unified the demographic (categorical) parameters with liver function (real value) parameters needed for consideration of their inter-relationships within directly clinically interpretable framework. Considering the established practice in HCC diagnostics [15,16], we determined ‘high’ and ‘low’ levels of each individual parameter using atercile-based dichotomization. For gender, reported alcoholism, evidence for hepatitis B and/or C and presence or absence of PVT the dichotomization was natural. For the other parameters, we tested several alternatives (50%:50%, quartiles) but found that tercile dichotomization with 2/3 of patients with the lowest parameter levels designated as “Low” phenotype and 1/3 of patients with the highest parameter levels designated as “High” phenotype was optimal for further processing. For age the “old” tercile was separated from the lower 2 “young” terciles by 55 years [17]. For the four significantly correlated parameter pairs, we used the two-thresholds that separate High from Low phenotypes, as shown in figure 1. This resulted in clinically familiar value cutoffs, such as bilirubin of 1.5 mg/dl, AST 200 IU/l and ALT 105 IU/l.

Step 3—Using actual data for each patient, an individual clinical profile was created by connecting all the actual parameter high, low, + and – levels (Figure 2) into a representation of their complete networked relationships. In figure 2 example, profiled patient is an older female, reporting alcoholism, diagnosed with HCV but not HBV, with AST<105 and ALT<80 IU/l, albumin >4.0 g/l, hemoglobin >15, bilirubin >1.5 mg/dl, INR >1.2, platelets >200 × 10⁹/l, AFP>29,000 ng/ml and presence of PVT. All these 4139 individual profiles were unified into a single schema, which carries new information about co-occurrence frequencies of all parameter levels (Figure S1).

Step 4—We found a simpler structure in the networked HCC clinical data for this cohort. The schema was completely decomposed into only 19 reference profiles C₁-C₁₉. These reference clinical profiles had to have identical co-occurrence frequencies between all the parameter levels. This ensured the independency of the results on the parameter ordering in the clinical profile: re-arranging the sections in figure 1 will generate identical data for subsequent steps. C₁-C₁₉ collects the information about the most frequent relationship co-occurrences of various parameter levels. C₁-C₁₉ thus serves as idealized clinical statuses (Figure S2).

Step 5—The 4139 individual profiles were then compared in turn to each of the 19 reference profiles and the total numbers (0–10) of mismatches in the relationships they describe were recorded as differences d_1 – d_{19} between the profiles.

Step 6—We next used logistic multiple regression [18] with variable selection algorithm (SigmaPlot11), using patient's 19 differences d_1 – d_{19} as independent variables, to predict whether an individual had a tumor mass (product of maximum tumor diameter and number of tumor nodules) smaller than 5.5 (1826 individuals, 44%) or larger (2313 subjects, 56%).

Results

Only the differences between patient actual clinical profiles and 9 reference clinical profiles out of 19 contributed significantly to the tumor mass classification. Of these, small differences (<6) from the 5 reference clinical profiles (C_1 , C_3 , C_6 , C_8 and C_{16}) resulted in high odds for a S-phenotype tumor mass and small differences (<6) from the 4 reference clinical profiles (C_5 , C_9 , C_{12} and C_{18}) resulted in the high odds for L-phenotype tumor mass. This logistic regression model correctly predicted 70% of the tumor mass categories in a 10-fold cross-validation (ROC area 0.78). We used the logistic regression equation to identify 2034 patients as S-subgroup and 2105 patients as L-subgroup. The distributions of tumor mass in these 2 subgroups had their means, (13.0 for L and 4.4 for S) significantly statistically different, $p=10^{-240}$, t -test. The significant differences were also seen systematically throughout the L and S tumor mass distributions. The Kaplan-Meier formalism (Figure 3) has shown with strong statistical significance that patients in the L-subgroup had a 2–4 fold greater odds of having a larger tumor. Equivalently, once the patient has been categorized in S or L-subgroup, then the odds of finding a given tumor mass were ~3 fold higher in the L-subgroup, compared with S-phenotype patients. This indicated that our findings are independent of specific choice of tumor mass threshold in optimization of the logistic regression classification model. The main result thus far was that liver function tests and patient demographic descriptors identified S and L-phenotypic groups with strongly statistically significant separation of their tumor mass distributions.

Logistic regression identified L-phenotype-associated reference profiles C_5 , C_9 , C_{12} and C_{18} , having in common high platelets/high AFP levels, accompanied by the presence of PVT and self-reported chronic alcohol consumption (Figure 4). The S-phenotype had 2 associated subgroups of reference clinical profiles: C_1 , C_3 and C_6 and C_8 and C_{16} . The former had in common: low platelets/low AFP and absence of PVT. The latter had in common low AST/low ALT, high albumin/high hemoglobin, low bilirubin/low INR and high platelets/high AFP.

L-phenotype associated reference profiles (C_5 , C_9) were male-related and C_{12} , C_{18} were female-related. C_9 described younger and C_5 described older (>55 years) patients. For the female-associated profiles, C_{12} described younger and C_{18} older patients. S-phenotype associated reference profiles C_6 (young) and C_{16} (older) were for female patients and C_1 (older), C_3 and C_8 (older) for male patients.

Trends between Individual Parameters and Tumor Mass in the S/L-Subgroups

The networked characteristic profiles for the L-subgroup are more homogeneous than those in the S-subgroup. We examined whether there were significant differences in typical parameter values for the same tumor mass that might be found in each of the S/L-subgroups.

We used a moving average filtering (Figure 5) where any tumor mass is characterized by the average of the clinical parameter values of 61 patients with the closest tumor masses [9]. We examined these trends in AFP (reflective of tumor growth) and platelet values and found

increasing AFP and platelet counts with increasing tumor mass in S- and L-subgroup with different rates and magnitudes (Figures 4). L-subgroup displayed a pattern of AFP/platelet level oscillations that were not observed in the S-subgroup. Importantly, these L-phenotype unique oscillations were characteristic for the same tumor masses in both AFP and platelet trends (Figures 5a and 5b).

The analysis of typical tumor-mass-related bilirubin level changes also showed differences in the 2 subgroups (Figure 5c). In the S-subgroup there was a shallow bilirubin increase as the tumor mass increased. In the L-subgroup, oscillations were found below tumor mass 20, which were not seen in the S-subgroup. The oscillations in bilirubin levels in the L-phenotype cohort occurred at the same tumor masses as those in AFP and platelet values in the L-subgroup. Additionally, there was a steady increase in bilirubin levels for increasing tumor mass beyond 20.

The mechanisms underlying the oscillations did not seem to have an obvious explanation from clinical practice. However, possible clues came from analysis of the number of tumor nodules typical for the given tumor mass (Figure 5d). We processed the data for tumor numbers in the same way as for other parameters and we found that oscillations in tumor numbers corresponded to spikes in the parameter trends, especially seen for tumor mass <30.

Examination of AST/ALT trends showed that they were steady in the S-subgroup and at higher levels than in the L-subgroup in the smallest tumors of equivalent mass <30 (Figure 6). In the L-subgroup, there was a steady increase in AST/ALT levels as tumor mass increased above 30.

Discussion

Our analysis used the results from a database constructed from HCC patients who were newly diagnosed by surveillance screening [9]. All parameters used were standard liver and blood count blood tests, which were already “processed” in the clinical practice. To find anything novel from these data, we needed to fundamentally modify the approach to re-analyzing these highly informative, but complex data in our meta-analysis. Instead of the standard liver test parameter values, we used relationships between the value levels of these parameters and considered any one of them in their complete context, in which every standard parameter level was evaluated in the full set of relationships to all other patient parameter levels. Instead of dealing with many complicated relationships between simple parameter values, we first transformed the raw liver test parameter data into differences between networked clinical profiles that contained, in manageable form, all the information about the clinical parameter interactions and parameter level relationships. We then used these transformed data as input into simple models separating two significantly different tumor phenotypes.

The simplicity of quantitative reconstruction of the real patient clinical outcome measure (tumor masses) from real-patient (input) data (which were personal differences in relationships between the real-world standard liver test data), allowed us to show that HCC is manageably heterogeneous, exhibiting two distinct series of liver test parameter relationship patterns for 2 very distinct subgroups of tumors.

Our finding that there were just two distinct subgroups of relationships between the real-world liver test parameter levels, each associated with a different tumor phenotype is not trivial or associated to arbitrary selection of just one tumor mass threshold for training classification. The significance of finding the 2 tumor mass distributions S and L with markedly different liver test parameter relationship characteristics distributions is primarily derived from $p=10^{-270}$ quantified statistical significance of differences in the S and L tumor

mass distribution means. Such extremely strongly significant separation of the outcome phenotype into two categories leaves little space for error and for having many other subgroups to consider.

Thus, the meta-analysis, which used just slightly more complicated data (parameter network graph distances) instead of just simple parameter levels, revealed a relative simplicity of the HCC overall patterns. The two tumor subgroups, though, were not entirely simply related to the liver test parameter relationships, but had manageable complexity. We found a total of 19 observable liver test parameter-level relationship types, for the extensive 4139 patient data. Out of those, 5 relationship types were associated in unity as a weighted-combination with one tumor phenotype (S), and 4 others were associated in unity as another weighted-combination with the second tumor type (L). Because the 5+4 reference clinical profiles were idealizations of the S and L clinical phenotype characteristics, an individual patient characterization involved a quantitative description of how close the actual pattern of relationships between parameter values were for an individual patient from all significant reference clinical profile patterns. In this approach, the single value of a parameter cannot change the classification. It was the majority of the parameter relationships matching the S or L-associated patterns that determined the classification.

Our main result is that we have shown that even with routine clinical patient parameter combinations, the added functionally relevant information about the disease can be extracted from trends and inter-dependencies of parameter values in a total parameter context. More importantly, we show that the complexity of HCC personalization is manageable and possible to treat in a few independent sub-dimensions of liver-test parameter level relationships, which however, must necessarily be treated together in the full context of all the other parameters of an individual patient.

We had 2 specific findings.

1. Figure 3 shows that there were always 2–4 times higher odds for larger tumors in L than in S phenotype group. The differences in the tumor mass trends in the two subgroups were significantly separated, showing the efficiency of our approach.
2. In the 2 phenotype groups, there was much greater homogeneity in the characteristic parameter patterns in L than in S. The rate of change for typical parameter values per unit change of tumor mass was always significantly higher for L-phenotype patients compared to S-phenotype patients, excepting the AST/ALT ratio, which was higher in S for tumor masses below 10 than for the same size tumors in L. One possible interpretation of these observations is that in S-phenotype patients, small tumors are associated with processes producing higher levels of the inflammatory markers, AST/ALT. We hypothesize that this might reflect the inter-connectedness of hepatic inflammation with tumor growth in the small tumors in this phenotype group. In L-subgroup, the simplest explanation of parameter levels oscillations might be consideration of the number of tumor nodules that composed the tumor mass. A relationship between platelet numbers and tumor size was recently reported [7]. Low platelets were interpreted to be a consequence of the portal hypertension that is secondary to liver fibrosis. We found that most small HCCs in 2 large western cohorts occurred in the presence of thrombocytopenia, whereas the largest tumors occurred in patients with significantly higher, but normal platelet values. By contrast, in the L-phenotype, the AST/ALT ratio only really increased as the tumor masses became quite large. This may reflect the parenchymal liver damage that occurs when a large tumor replaces underlying liver. Also in the L-phenotype, but not in S, several additional liver parameters showed oscillations in their typical values, as the tumor mass increased.

We found a relationship between these oscillations and the numbers of tumors (Figure 5).

Given the lesser association of changes in inflammatory markers in the L-phenotype, we consider that other factors, likely tumor-related, may be more important on the growth of these tumors. Such factors likely include genetic drivers of HCC cell growth. In the L-phenotype, the observed higher levels of various parameter contributions from multiple nodules to the total parameter levels could be additive.

Conclusions

The recognition that patients with newly diagnosed HCC can be identified with either of 2 different phenotypic subgroups based on common clinical parameters, each subgroup having distinctive biological characteristics, provides a training set to be used for future validation of its clinical and prognostic usefulness.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

PP was supported in part by ERC-CZ LL1201 program CORES and by NIH grant CA 82723 (BC).

References

1. Okuda K, Nakashima T, Kojiro M, Kondo Y, Wada K. Hepatocellular carcinoma without cirrhosis in Japanese patients. *Gastroenterology*. 1989; 97:140–146. [PubMed: 2542116]
2. Venook AP, Papandreou C, Furuse J, de Guevara LL. The incidence and epidemiology of hepatocellular carcinoma: a global and regional perspective. *Oncologist*. 2010; 15:5–13. [PubMed: 21115576]
3. Kumada T, Toyoda H, Kiriya S, Sone Y, Tanikawa M, et al. Incidence of hepatocellular carcinoma in patients with chronic hepatitis B virus infection who have normal alanine aminotransferase values. *J Med Virol*. 2010; 82:539–545. [PubMed: 20166172]
4. Trevisani F, D'Intino PE, Caraceni P, Pizzo M, Stefanini GF, et al. Etiologic factors and clinical presentation of hepatocellular carcinoma. Differences between cirrhotic and noncirrhotic Italian patients. *Cancer*. 1995; 75:2220–2232. [PubMed: 7536121]
5. Rosa JC, Chaves P, de Almeida JM, Soares J. Hepatocellular carcinoma. Rare forms of presentation. *Acta Med Port*. 1995; 8:243–245. [PubMed: 7625220]
6. Lok AS, Seeff LB, Morgan TR, di Bisceglie AM, Sterling RK, et al. Incidence of hepatocellular carcinoma and associated risk factors in hepatitis C-related advanced liver disease. *Gastroenterology*. 2009; 136:138–148. [PubMed: 18848939]
7. Carr BI, Guerra V, Pancoska P. Thrombocytopenia in relation to tumor size in patients with hepatocellular carcinoma. *Oncology*. 2012; 83:339–345. [PubMed: 23006937]
8. Carr BI, Guerra V, De Giorgio M, Fagioli S, Pancoska P. Small hepatocellular carcinomas and thrombocytopenia. *Oncology*. 2012; 83:331–338. [PubMed: 23006906]
9. Lu SN, Wang JH, Liu SL, Hung CH, Chen CH, et al. Thrombocytopenia as a surrogate for cirrhosis and a marker for the identification of patients at high-risk for hepatocellular carcinoma. *Cancer*. 2006; 107:2212–2222. [PubMed: 17019738]
10. Pancoska P, Carr BI, Branch RA. Network-based analysis of survival for unresectable hepatocellular carcinoma. *Semin Oncol*. 2010; 37:170–181. [PubMed: 20494709]
11. Hoshida Y, Villanueva A, Kobayashi M, Peix J, Chiang DY, et al. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N Engl J Med*. 2008; 359:1995–2004. [PubMed: 18923165]

12. Leonardi GC, Candido S, Cervello M, Nicolosi D, Raiti F, et al. The tumor microenvironment in hepatocellular carcinoma (review). *Int J Oncol*. 2012; 40:1733–1747. [PubMed: 22447316]
13. Utsunomiya T, Shimada M, Imura S, Morine Y, Ikemoto T, et al. Molecular signatures of noncancerous liver tissue can predict the risk for late recurrence of hepatocellular carcinoma. *J Gastroenterol*. 2010; 45:146–152. [PubMed: 19997856]
14. Diestel, R. *Graph theory*. 4. Springer; Heidelberg, New York: 2010.
15. Yang JD, Sun Z, Hu C, Lai J, Dove R, et al. Sulfatase 1 and sulfatase 2 in hepatocellular carcinoma: associated signaling pathways, tumor phenotypes, and survival. *Genes Chromosomes Cancer*. 2011; 50:122–135. [PubMed: 21104785]
16. Tanaka K, Sakai H, Hashizume M, Hirohata T. Serum testosterone:estradiol ratio and the development of hepatocellular carcinoma among male cirrhotic patients. *Cancer Res*. 2000; 60:5106–5110. [PubMed: 11016636]
17. Carr BI, Pancoska P, Branch RA. HCC in older patients. *Dig Dis Sci*. 2010; 55:3584–3590. [PubMed: 20238246]
18. Hall, M.; Eibe, F.; Holmes, G.; Pfahringer, B.; Reutemann, P., et al. *SIGKDD Explorations*. 2009. The WEKA Data Mining Software: An Update; p. 11

Figure 1.

Set of four parameter pairs that have the highly correlated trends. One point represents a patient, in gray we show the upper tercile of patients identified with "high" levels of both parameters. In black are 2/3 of patients with "low" parameter levels. The boundary between the high and low levels is defined by the two threshold values indicated in every picture.

Figure 2.

Example of 10-partite individual clinical profile of a patient. F=female, M=male, O=age>55, Y age<55 years, +/- presence/absence of indicated parameter. H/L - correlated parameter pair levels, shown in figure 1. The patient's clinical profile is recovered from this scheme by following the black line. BILI = bilirubin, Alb=albumin, Hemo=hemoglobin

Figure 3. Modified Kaplan-Meier characterization of odds for a given the tumor mass in S- and L-subgroups. The dotted lines are 95% confidence intervals for respective odds curves. Double arrow in a) is 3-fold odds difference. Note different scales for the largest tumor masses in b).

Figure 4.

Top two panels: columns P1-P10 correspond to 10 parts of the clinical profile shown in Fig. 2. The two sub-columns indicate one of the two levels for the respective parameters. The actual levels for given reference clinical profiles are shown by "1" in black field. Left columns: Reference profile ID and coefficient in the classification logistic regression model. Top: reference clinical profiles associated to L-subgroup, bottom: reference clinical profiles associated to S-subgroup. Bottom two panels: percentage of commonality of all reference clinical profile levels in respective sections P1-P10.

Figure 5.

Typical parameter levels as function of tumor mass in S- (gray) and L-subgroups (black). Typical values are results of moving average processing (see Methods) a) AFP, b) platelets, c) bilirubin, d) number of tumors.

Figure 6.

Typical levels of AST/ALT ratio for S- and L-subgroups, rank-ordered by tumor mass. a) First 600 patients with tumor masses 1–2.5 in S-subgroup and 1–7 in L-subgroup.

b) Remaining patients with larger tumor masses. Gray – S-subgroup, black – L-subgroup

Table 1

Demographic and clinical characterization of patients.

		Number	Percent in study	Number	Percent in study	
Gender	female	1033	24.9%	male	3106	75.1%
Age	Younger<55 year	1432	34.6%	Older>55 year	2707	65.4%
Alcohol	[-]	2950	71.2%	[+]	1189	28.7%
HBV	[-]	2010	48.6%	[+]	2129	51.4%
HCV	[-]	2520	60.9%	[+]	1619	39.1%
PVT	[-]	3187	77%	[+]	952	23%
AFP	low<200	2671	64.5%	high>200	1468	35.5%
Bilirubin	low<1.2	2618	63.2%	high>1.2	1521	36.8%
ALT	low<40	1693	40.9%	high>40	2476	59.1%
AST/ALT	low<1.0	1344	32.5%	high>1.0	2795	67.5%
albumin	low<3.0	951	23.0%	high>3.0	3188	77.0%
hemoglobin	low<13.0	2236	54.0%	high>13.0	1903	46.0%
platelets	Low 125	1603	38.7%	high>125	2536	61.3%
INR	low<1.0	1355	32.7%	high>1.0	2784	67.3%

PVT, hepatitis B and C antigens (HBV, HCV) absence/presence and self-not/reported alcoholism are shown as [-]/[+]. Concentration units are: albumin [g/dL], hemoglobin [g/dL], bilirubin [mg/dL], platelet [10^3 /dL], AFP [ng/dL], INR [r.u.], AST [U/L], ALT [U/L].

Table 2

Reference profile ID and coefficient in the classification logistic regression model.

Number of tumors	Patients/percent of cohort	Size range [cm]	Mean [cm]	Median [cm]	std. deviation [cm]
1	2147 (51.9%)	1–27.7	4.9	3.2	4.1
2	575 (13.9%)	1–22.2	4.4	3.3	3.3
3	178 (4.3%)	1–15	3.9	3.0	2.5
>3	1239 (30%)	0.9–26.0	7.9	8.5	4.3