Validation of common typical time of disease duration for HCC patients using the Fisher information processing of tumor imaging results combined with network phenotyping strategy quantification of individual patient clinical profile patterns.

Petr Pančoška[a,b], Lubomír Skála[c†], Jaroslav Nešetřil[b], Brian I. Carr [d]

a) University of Pittsburgh
b) Institute of Theoretical Informatics, Charles University, Prague,
c) Department of Chemical Physics and Optics, Faculty of Mathematics and Physics, Charles University Prague.
d) Department of Gastroenterology and Liver Diseases, Tel-Aviv, Sourasky Medical Center, Tel Aviv, Israel.

Grant support: ERZ-CZ LL1201 (CORES) to P.P. and J.N.

† Unfortunately, Professor Lubomír Skála passed away on May 8, 2015.

Abbreviations: HCC: hepatocellular carcinoma; Tmass: product of the number of tumor nodules and the maximal diameter of the tumor; ln(Tmass): natural logarithm of tumor mass; ln(ρ(Tmass)): natural logarithm of the probability density ρ of finding given Tmass; Tonset: time to disease onset, number of days from the tumor appearance to the baseline screening, NPS: network phenotyping strategy; $\delta(\boldsymbol{P_i}, \boldsymbol{HL_1})$: number of differences in clinical parameter relationship graph $\boldsymbol{P_i}$, observed for the patient ***i*** and graph and $\boldsymbol{HL_1}$, representing the comprehensive pattern of relationships between normal clinical parameter values; TDD: total disease duration; OVS: overall survival.

Conflicts of interest: none, by any author

E-mail: brianicarr@hotmail.com

ABSTRACT

A primary goal of current clinical cancer research is the identification of prognostic tumor subtypes. It is increasingly clear that tumor growth depends on both internal tumor factors, and factors that are external to the tumor, such as microenvironment. We recently showed that parameter values alone are less important than the patterns of all patient parameters together for the identification of prognostic subtypes and have identified a network phenotyping strategy method to quantitatively describe the dependency of the tumor on the environment, to characterize HCC subtypes. We have also shown that information about tumor mass together with patterns of other prognostic factors is related to survival. We now use a different patient cohort to validate this prognostic approach. A main finding is our identification of a common time of total disease duration (TDD) for every HCC patient. Clinical prognosis at the time of baseline patient evaluation is then calculable as the difference between TDD and the time from disease onset to diagnosis (Tonset). We show that the total pattern of all parameter values and the differences in the relationships between this pattern and a reference pattern that, together with the tumor mass, best reflects the patient prognosis at baseline. Our approach led us to identify 15 different composite HCC subtypes. Our results highlight the nearly identical TDD in all patients, which must therefore be a characteristic of the HCC disease, as opposed to the variable quantity of Tonset, which is impacted by multiple macro- and micro-environmental factors.

INTRODUCTION

It has been recently appreciated that tumors are not completely independent (oncogene-driven) from their environment, but their growth can be in part explained by signals from their macroenvironment (sex hormones, nutrition) and microenvironment (growth factors, inflammatory cytokines, cellular milieu) (1-4). This applies also to hepatocellular carcinomas or HCCs (5-14).

We recently found how to quantitatively describe this dependency on the environment and characterize HCC subtypes (S and L) using new information that is available in the standard screening data once all data from a patient are

kept coherently throughout the processing (15).

We previously concentrated on explaining Tmass (a product of maximum tumor diameter and number of tumor nodules), as a disease end point or outcome, but not used it for prognosis. With this separation of information about tumor from the information about the other clinical parameters, we have shown that Tmass and its associated clinical factors are both micro-environmental and macro-environmental.

In the last paper, we put information about the tumor mass (Tmass) back together with other the prognostic factors and derived how this combined information is related to survival. In order to use the new personally coherent part of information in the standard data, a personal characterization of tumor growth was combined with personal coherent status of an individual patient's clinical characteristics to estimate survival (prognosticate). This goal required relating tumor mass to time of tumor growth. We found (16) that to be able to do that, it is sufficient to require that the information about tumor mass, an observation from the radiology scan (CAT or MRI scan), has to be equal to the information about the sum of the tumorigenesis processes (tumor biology, reflecting internal tumor together with external patient influences). We then derived from this condition the relationship of tumor mass in an individual patient to time of tumor origin or onset until time to diagnosis using Fisher information formalism (16).

We now use a different patient cohort to validate our recent results and we give some clinical patient examples of the consequences of this approach.

METHODS

We analyzed prospectively-collected data in the Italian Liver Cancer (ITA.LI.CA) study group database of HCC patients accrued at 11 centers. Complete data, allowing us to construct personal clinical patterns using the Network Phenotyping Strategy (NPS) were available for N=1980 patients (17) and the database management conformed to Italian legislation on privacy. This study conforms to the ethical guidelines of the Declaration of Helsinki. Approval for the study on de-identified patients was obtained by the Institutional Review Board of participating centers. The clinical parameter data were processed exactly as previously (16).

RESULTS

Fig. 1 shows the conversion of Tmass into the first parameter needed to determine T_{onset} (time from tumor onset till clinical diagnosis). It represents a standard histogram of tumor masses, normalized to unit area.

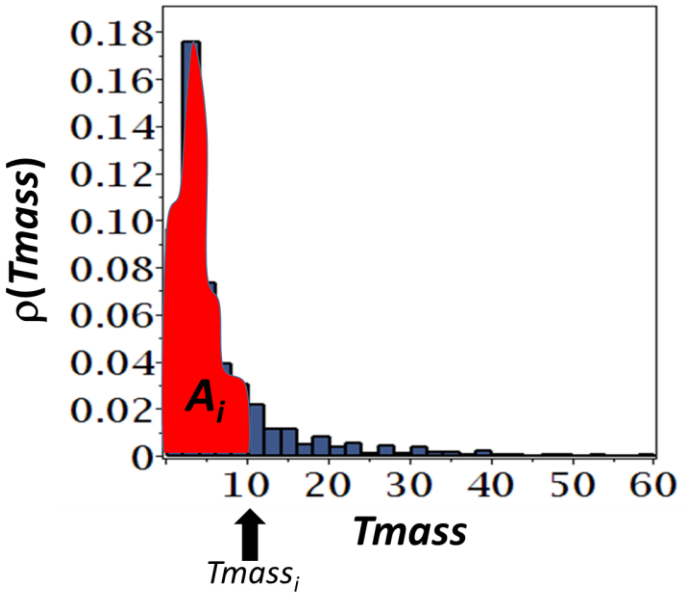

Fig.1 Explanation of extracting the information about personal component $\boldsymbol{A_i}$ (red area) of ***Tonset*** for patient with $\boldsymbol{Tmass_i}$=10 from the histogram of tumor masses in this study (blue bars), normalized to represent the $\rho(Tmass)$. Natural cubic spline interpolation and integration was used to obtain the numerical values of $\boldsymbol{A_i}$.

We previously showed that the first parameter that is needed for computing T_{onset} is the integrated area of this histogram for Tmass ranging between zero and an individual patient tumor mass. This is shown by the red area for a

patient with tumor mass 10 in Fig. 1. For the current cohort that is based predominantly on diagnosis through surveillance of patients at risk of HCC, relatively to the previous cohort of randomly diagnosed patients (16), we did the same testing of the normality of the Tmass distribution and found it significantly non Gaussian as was found previously for US patients. This in clinical sense means that through applying Fisher information formalism, we are using in our analysis the information about tumorigenesis factors, which made the histogram of the tumor masses non-Gaussian to estimate the first component of Tonset.

We have previously shown by mathematical argument that requiring the equality between the information about the tumor contained in the MRI scan and the information about the various internal, biological processes determining the tumorigenesis leads to power-law of tumor growth. Having this power-law formula permits us to trace the tumor mass growth back in time from the moment of clinical screening to the disease onset. We can therefore apply a simple test if the tumor masses, actually observed for the patients in any cohort, are growing according to the law, derived by this approach. The mathematical form of the growth law predicts that the log of normalized histogram intensities is linearly related to the log of the observed tumor mass. This linear relationship was indeed found in the previous paper (16) as well as for the current cohort (Fig. 2).

Moreover, the current (validation) and previous (original) patient cohort exhibit quantitatively similar log-linear relationships between the tumor mass and histogram intensities for most of the patient cohort. The only difference between the previously published and current cohort is the presence of a minor group of patients in the former cohort, with very large, slower-growing tumors, that are dependent on more complex tumor-generating processes. Possibly, this relates to the clinical randomness of

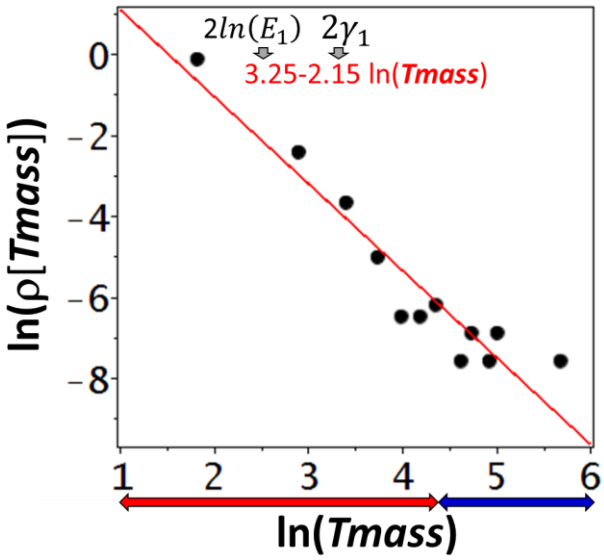

Fig.2 . Explanation of extracting the information about HCC-specific parameters of ***Tonset*** from tumor growth law, which indicated that log-log transformed ***Tmass*** histogram should be represented by linear functions. The line was least-square fitted, resulting in the numerical value for E_j and γ_j as is shown.

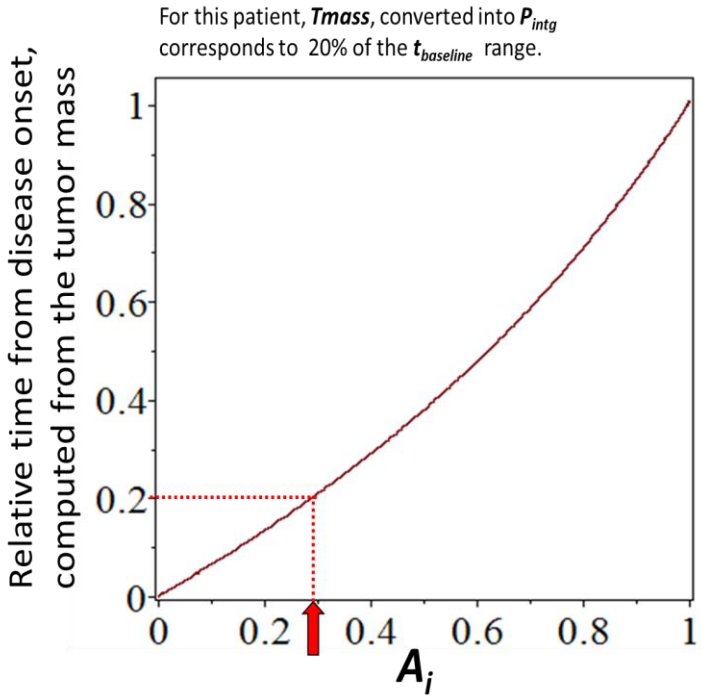

Fig.3 Plot of the dependence of T_{onset} on A_i, as derived from Fisher information method. We show by example that a patient with the integrated area of 0.3 in the Fig. 1 have the T_{onset} in the 20% of the total range observed.

diagnosis in the previous cohort, whereas a large percent of the current patients were diagnosed through screening.

Fig. 3 shows the plot of the tumor mass related parameter from each patient, defined by the personal integrated area, shown by example in Fig. 1, in relationship to the time to disease onset in relative units.

This relationship is qualitatively and quantitatively identical to the previously published cohort, excepting the absence of the small component, characterizing patients with very large and slower growing tumor masses that was previously noted.

Fig 4 describes the optimization-based determination of the last parameter in the tumor-growth law, which was done as described in details in ref. 16. The result of the optimization

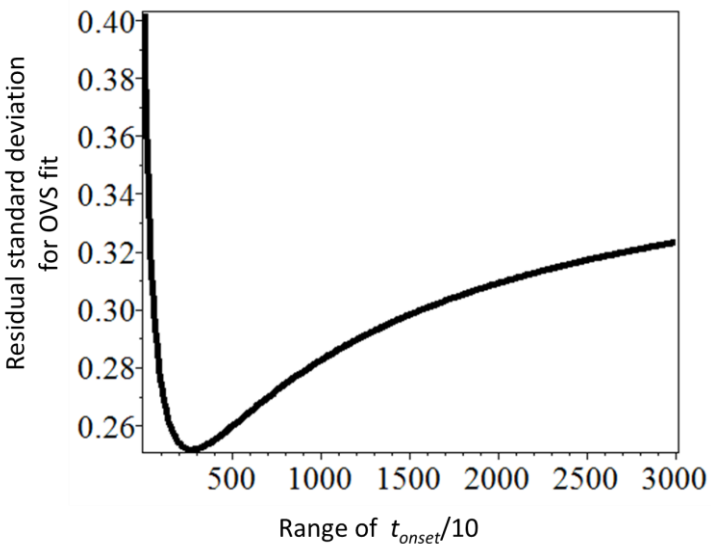

Fig.4 Plot of the residual standard deviations of $\boldsymbol{OVS_C}$ – $\delta(P_i,HL_1)$ fits for systematically varied values of $\boldsymbol{T_{max}}$. The optimal value is in the minimum of this curve.

procedure demonstrates existence of a single common constant, converting the time to disease onset from relative units to actual days. This result again confirms the existence and uniqueness of the characteristic optimal value of this constant, found in the previously published cohort.

We have also previously shown (16) that the tumor growth rates are related to the personal clinical patterns of individual HCC patients. We use k-partite graphs to capture quantitatively these patient-unique contributions of clinical parameter values and the pattern of relationships between the values as they were simultaneously found for a patient, within the framework of a Network Phenotyping Method (NPS) (15).. For personal characterization of the clinical status pattern in NPS we quantify the differences between patient composite clinical profiles. This quantification of the differences between the individual patterns was done systematically, in terms of $\delta(\boldsymbol{P_i},\boldsymbol{HL_1})$ values, which resulted from comparing patient's composite clinical profiles $\boldsymbol{P_i}$ to a common profile $\boldsymbol{HL_1}$ that represented the best possible prognosis based upon all the considered parameters being normal and counting the differences between them. The Fig 5 shows that there is a linear relationship

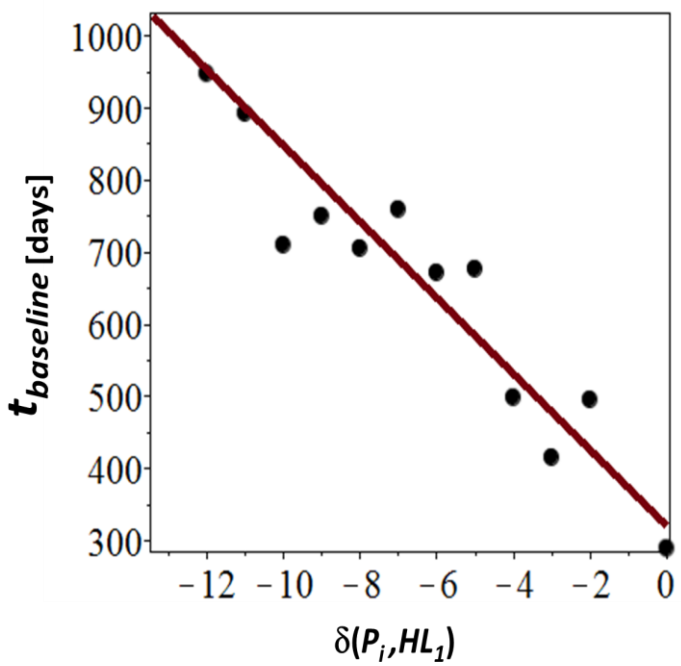

Fig.5 Linear relationship between the clinical profile differences and mean values <***Tonset***> computed as average of individual values of ***Tonset*** for each patient HCC subgroup.

between T_{onset} and these personal composite pattern differences $\delta(\boldsymbol{P_i,HL_1})$ as we previously described (16). This validates our previous finding (16) that tumor mass alone does not provide full prognosis of the survival and needs to be considered in the full context of patient individual clinical profile.

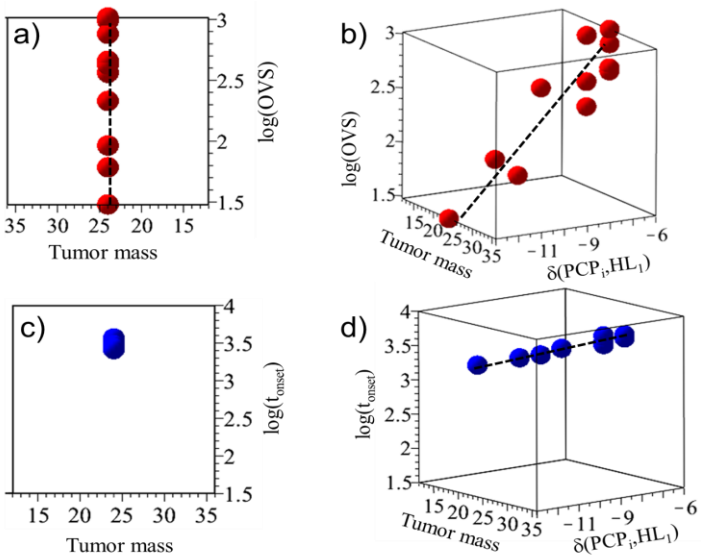

Fig.6 a) Example of survival variability for patients with the same Tmass. b) 3D plot of the dependence of OVS for patients with the same Tmass on the $\delta(P_i,HL_1)$, showing the linear relationship. c) After the correction of OVS for $\boldsymbol{T_{onset}}$, all patients exhibit the same TDD. d) 3D plot of the identity in TDD after the $\delta(P_i,HL_1)$-dependent correction of OVS by T_{onset}, into TDD.

Fig. 6 clarifies this general relationship for a simpler example of 10 different patients who have an identical tumor mass of 25 (Fig. 6a). It is clear from this figure that tumor mass alone cannot predict the observed 2 log differences in the survivals amongst these 10 patients.

However, there is a direct proportionality between the patient's clinical pattern profile difference from normal pattern $\boldsymbol{HL_1}$ and an individual patient survival (Fig. 6b). When this clinical profile based ordering is taken into consideration for each of the 10 different patients with identical Tmass, Fig. 6c and 6d show that after correction, the time of disease duration (TDD) is the same for all patients. This conclusion of a commonality of TDD for HCC patients is valid for all tumor masses (Fig. 7).

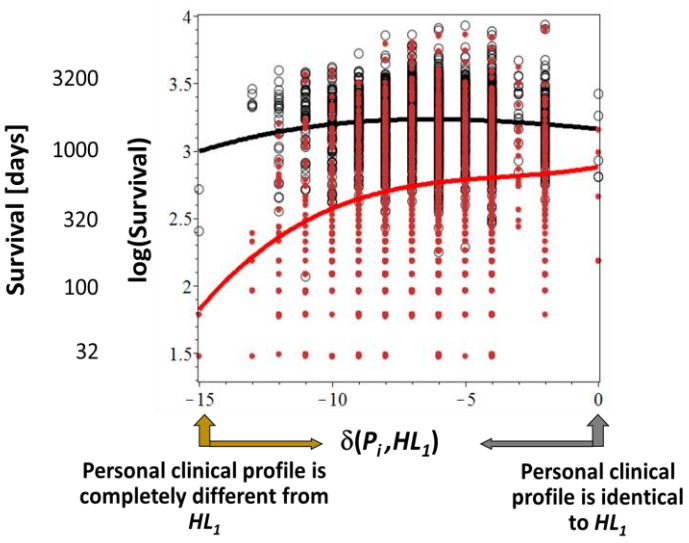

Fig.7) Relationships between $\delta(P_i,HL_1)$, ***OVS*** (black points) and $\boldsymbol{OVS_C}$ (blue circles). The lines are least-square fits of the relationships by cubic model. 15 groups of patients with matching differences $\delta(P_{iK},HL_1)$of their personal clinical relationship profiles form the vertical groups of points.

Thus, differences in individual patient prognosis relate to differences in Tonset i.e. the time in the course after tumor initiation in the natural history at which the HCC is diagnosed. It was considered as intractable, but we have shown that firstly, after the tumor-mass and NPS-based correction, there is a common overall duration of the disease, or TDD; secondly, that we can estimate time to onset (or T_{onset}) from the information in the tumor mass and the composite clinical pattern, using standard clinical parameters and thirdly, from these results we can estimate the actual clinical overall survival (OVS) from the difference between the common clinical duration and the personal time to disease onset or T_{onset} (till diagnosis from disease onset): OVS = TDD - T_{onset}.

DISCUSSION

Knowledge of the time from initiation of a tumor till its diagnosis (time to onset) appears to be a problem, because, at the time of patient's initial visit, it is seemingly random. Thus, it has been assumed that to estimate this Tonset, one needs to know the size of the tumor at 2 or more different time points, to determine the growth law together with the rate of growth and then back calculate the time when the tumor mass was zero. We have shown that, using the Fisher information approach (16), there is an alternative way to this calculation, which uses the shape of the non-random distribution of the actual tumor masses and then determine the tumor growth law from the requirement of equality between information about the biological tumorigenesis factors and information collected by tumor imaging. For the results to be practical for translational clinical application, it would be best if there is a common (typical) time of patient disease duration (TDD). We therefore formulated this restriction on TDD as our initial hypothesis and validated it by the full compliance of the results which we obtained with the derived tumor growth law properties. This alternative approach substitutes the determination of tumor growth law from longitudinal observations by using the fact that in this cohort, patients provided 1980 time of growth points for this analysis, i.e. every patient in the cohort represents a different time of tumor growth, when presenting at initial clinical diagnosis. This allowed us to estimate the parameters of the tumor growth model, derived to satisfy the maximal compliance between the tumor growth factors and the tumor pathology at baseline. This then allowed us to calculate Tonset. The result of this computation not only provided a Tonset for each patient, but also showed that the mathematical formulation of the tumor growth law is fully compatible with the imaging data that are reflected in Tmass (Fig. 2), and thus validated also the initial assumption of a common time of total disease duration for every HCC patient (Fig. 7).

However, knowledge of Tmass-based Tonset is only one important factor here. Other patient characteristics, such as gender (hormonal influences), age and microenvironmental factors such as inflammation, are also important factors that impact tumor growth (18, 19). We found here and previously that to quantify the impact of these additional factors upon the clinical behavior of the HCC, we needed more than just their clinical values. The clinical characterization of these factors was found to be related to the composite pattern of the standard screening clinical characteristics, observed simultaneously for every patient. We also needed the quantitative descriptor of the inter-patient differences $\delta(\boldsymbol{P_i,HL_1})$ in these patterns, which is to be used as input for diagnostic and prognostic decisions. We quantified these differences simply in terms of the number of differences in the pattern relationships, when a patient's clinical pattern is compared to a baseline reference pattern $\boldsymbol{HL_1}$ (16).

The need for considering not only the tumor mass, but also the pattern, reflecting the patient's actual clinical status, is exemplified in Fig. 6, which shows 10 different patients, who had identical tumor masses, but had quite different survival from each other, because they had 10 different patterns of their total parameters. The most important result of our new analytical paradigm applied to HCC is that there is a definite proportionality between the survival and the inter-patient differences of their composite characteristic clinical patterns at baseline diagnosis. Our overall results have shown (see Fig. 7) that this (linear) relationship between survival and pattern differences are found for all tumor masses observed in the two cohorts, but was not observed when only the actual clinical values of any considered parameter such as AFP or bilirubin were tested for this relationship. Rather, it is the total pattern of all the parameter values and the differences in the relationships between these total patterns and the reference pattern that, together with Tmass, best reflects

the patient prognosis at baseline. This seems to us to indicate that all the parameters that we routinely measure in clinical practice are not stand-alone in their significance, but appear to interact through the internal functional disease-related process with all the other parameters. In addition, we believe our results show that it is important to consider these interaction patterns in a strictly personalized way for each patient, because without this relationship pattern of parameters that are unique for each patient, we cannot obtain this clear relationship to survival.

Together, the clinical prognosis at the time of baseline patient evaluation is then calculable as OVS = TDD - Tonset (total disease duration minus the time from disease onset to diagnosis). The Excel worksheet implementation of the computational algorithm, obtained as result of our work, can be found and downloaded as Supplementary Material to Ref 16. In that tool, the user enters the raw clinical data and the worksheet automatically computes the prognosis.

The final point to emphasize is the identification of the source of the new information, which led to our findings. Conventional clinical analysis typically considers a limited number of parameters in relation to an end-point such as tumor size or patient survival. This is typically done using standard statistical techniques, which consider each parameter individually and independently. This means that patient's information is considered via separate parts of the total clinical picture. If interacting parameters have to be considered, incorporating them fully into the calculations lead to prohibitive complexity and need for large number of patients in a study to maintain sufficient power. Thus, the interaction considerations have frequently been restricted to a few parameters such as platelets, where the thrombocytopenia that reflects the fibrotic process of cirrhosis, was recently shown to be related to small size of HCC (20) and by contrast, massive size HCCs often have normal platelet counts or even thrombocytosis (21).

The standard analytic approach thus destroys information about the total coherence of the clinical parameter values, which in turn characterizes the personal uniqueness of the clinical pattern for each patient. By contrast, the Network Phenotyping Strategy approach presented here, describes quantitatively and in directly clinically interpretable form a total interacting system that reflects both parameter values, relationships between values and their individual coherence for every patient.

The NPS-based approach described here led us to identify 15 different composite HCC subtypes, each clinically characterized by increasing difference of the patient pattern from normal pattern. The major finding was the linearity of relationship of these quantitative clinical pattern characteristics to Tonset and OVS for the patients with their personal clinical statuses in between extremes of normal versus maximally different baseline clinical patterns (Fig. 5). This ordering of the clinical statuses of patients provides a predictable correction of the OVS, leading to the commonality of the total disease duration (TDD) of all patients in all 15 groups (see horizontal line in Fig. 7). These results also highlight the nearly identical TDD in all patients, which must therefore be a characteristic of the HCC disease in people, as opposed to the variable quantity of Tonset, which is impacted by multiple macro- and micro-environmental factors.

Throughout the time of disease onset, this common TDD is "spanned" by spectrum of slowly growing to aggressive tumors. We understand a slowly growing tumor as one in which there is small or no growth of the tumor size on the scan at 2 different time intervals. By contrast, aggressive tumors are considered to be those which increase in size within the same time, often associated with constitutional symptoms and worsening liver function parameters, which reflect the increased liver damage associated with an enlarging tumor. Our results show (Fig.7) that there is a continuum of tumor states between these two extremes (benign is at $\delta(\boldsymbol{P_i,HL_1})=0$, aggressive at $\delta(\boldsymbol{P_i,HL_1})=-15$). Thus, each individual patient has his/her characteristic,

variable Tonset, which in its value integrates contributions from both Tmass and from the individual pattern of clinical characteristics. When this characterization of the personal variability of the disease progression, which in turn describes the degree of aggressiveness of the individual patient's tumor, is combined with the finding that there is a constant TDD, then OVS can be then more reliably estimated.

REFERENCES

1. Verloes R, Kanarek L. Tumour microenvironment studies open new perspectives for immunotherapy.Arch Int Physiol Biochim. 1976;84:420-2
2. Tarin D, Price JE. Influence of microenvironment and vascular anatomy on "metastatic" colonization potential of mammary tumors. Cancer Res. 1981;41:3604-9.
3. Schirrmacher V. Shifts in tumor cell phenotypes induced by signals from the microenvironment. Relevance for the immunobiology of cancer metastasis. Immunobiology. 1980;157:89-98
4. Radinsky R, Fidler IJ. Regulation of tumor cell growth at organ-specific metastases.In Vivo. 1992;6(4):325-31. Review
5. Zheng XY, Ling ZY, Tang ZY, Liu YK, Feng XL, Zhuang W. The abundance of NM23-H1 mRNA is related with in situ microenvironment and intrahepatic metastasis in hepatocellular carcinoma. J Exp Clin Cancer Res. 1998;17:337-41
6. Giannelli G, Bergamini C, Fransvea E, Sgarra C, Antonaci S. Laminin-5 with transforming growth factor-beta1 induces epithelial to mesenchymal transition in hepatocellular carcinoma. Gastroenterology. 2005;129:1375-83.
7. Budhu A, Forgues M, Ye QH, Jia HL, He P, Zanetti KA, Kammula US, Chen Y, Qin LX, Tang ZY, Wang XW.
Prediction of venous metastases, recurrence, and prognosis in hepatocellular carcinoma based on a unique immune response signature of the liver microenvironment. Cancer Cell. 2006;10:99-111
8. Sagmeister S, Eisenbauer M, Pirker C, Mohr T, Holzmann K, Zwickl H, Bichler C, Kandioler D, Wrba F, Mikulits W, Gerner C, Shehata M, Majdic O, Streubel B, Berger W, Micksche M, Zatloukal K, Schulte-Hermann R, Grasl-Kraupp B. New cellular tools reveal complex epithelial-mesenchymal interactions in hepatocarcinogenesis. Br J Cancer. 2008;99:151-9
9. Tu T, Budzinska MA, Maczurek AE, Cheng R, Di Bartolomeo A, Warner FJ, McCaughan GW, McLennan SV8 Shackel NA. Novel aspects of the liver microenvironment in hepatocellular carcinoma pathogenesis and development. Int J Mol Sci. 2014;15:9422-58.
10. Hernandez-Gea V, Toffanin S, Friedman SL, Llovet JM. Role of the microenvironment in the pathogenesis and treatment of hepatocellular carcinoma. Gastroenterology. 2013;144:512-27.
11. Utsunomiya T1, Shimada M, Imura S, Morine Y, Ikemoto T, Mori M.Molecular signatures of noncancerous liver tissue can predict the risk for late recurrence of hepatocellular carcinoma. J Gastroenterol. 2010;45:146-52.
12. Hoshida Y, Villanueva A, Kobayashi M, Peix J, Chiang DY, Camargo A, Gupta S, Moore J, Wrobel MJ et al. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. N Engl J Med. 2008;359:1995-2004.
13. Carr BI, Guerra V. HCC and its microenvironment. Hepatogastroenterology. 2013;60:1433-7.
14. Pancoska P, Carr BI. Macro- and micro-environmental factors in clinical hepatocellular cancer. Semin Oncol. 2014;41:185-94
15. Pancoska P, Lu SN, Carr BI. Phenotypic Categorization and Profiles of Small and Large Hepatocellular Carcinomas. J Gastrointest Dig Syst. 2013;Suppl 12. pii: 001.
16. Pančoška P, Skála L, Nešetřil J, Carr BI. Evaluation of Total Hepatocellular Cancer Lifespan, Including Both Clinically Evident and Preclinical Development, Using Combined Network Phenotyping Strategy and Fisher Information Analysis. Semin Oncol. 2015;42:339-346
17. Carr BI, Pancoska P, Giannini EG, Farinati F, Ciccarese F, Ludovico Rapaccini G, Di Marco M, Benvegnù L et al; Italian Liver Cancer Group. Identification of two clinical hepatocellular carcinoma patient phenotypes from results of standard screening parameters. Semin Oncol. 2014;41:406-14
18. Carr BI, Pancoska P, Branch RA. Tumor and liver determinants of prognosis in unresectable hepatocellular carcinoma: a large case cohort study. Hepatol Int. 2009;4:396-405
19. Horino K, Beppu T, Kuroki H, Mima K, Okabe H, Nakahara O, Ikuta Y, Chikamoto A, Ishiko T et al.
Glasgow Prognostic Score as a useful prognostic factor after hepatectomy for hepatocellular carcinoma.
Int J Clin Oncol. 2013;18:829-38.
20. Carr BI, Guerra V, De Giorgio M, Fagiuoli S, Pancoska P. Small hepatocellular carcinomas and thrombocytopenia.Oncology. 2012;83:331-8
21. Carr BI, Guerra V. Features of massive hepatocellular carcinomas.Eur J Gastroenterol Hepatol. 2014;26:101-8