Published in final edited form as:

J Inonu Liver Transplant Inst. 2024 December ; 2(3): 109–116. doi:10.14744/jilti.2024.63935.

A New Approach to Analysis of Clinical Data and Prognostication for Patients with Hepatocellular Carcinoma, Based Upon a Network Phenotyping Strategy (NPS) Computational Method

Brian Carr[1], **Patricia Sotákov**[2], **Petr Pancoska**[3]

[1]Liver Transplantation, Inonu University, Malatya, Türkiye

[2]Institute of Theoretical Informatics, Charles University, Prague, Czech Republic

[3]Faculty of Sciences, Charles University, Prague, Czech Republic

Abstract

Objectives: There is a multi-component nature of the influences on HCC progression but integrating them has been difficult. Network phenotyping strategy (NPS) integrates all multi-component relationship facets of HCC progression and aims to lead to a new way of understanding human HCC biology.

Methods: We converted baseline patient demographics, tumor characteristics, blood hematology and liver function test results, consisting of values of 17 standard clinical variables, collected time-coherently at the index visit, into a graph-theoretical data representation.

Results: These data were analyzed by NPS, which processes the patient parameter values together with their complete relationships network. NPS identified 25 disease-progression ordered HCC phenotypes. Clinically relevant NPS results are a) Portal vein thrombosis incidence during HCC progression stratified into 5 narrow ranges; b) NPS identified patients according to aggressive, slow and intermediate tumor growth sub-types; c) Personalized prognostication of mortality was achieved by the 25 NPS phenotypes, independently optimized for respective phenotype sub-cohorts.

Conclusion: The NPS results were implemented as an internet application (https://apkatos.github.io/webpage_nps), where input of 17 clinical parameters provides the patient phenotype, phenotype-characteristic average mortality and personal survival estimate.

Address for correspondence: Brian Carr, MD. Liver Transplantation, Inonu University, Malatya, Türkiye, brianicarr@hotmail.com.

Disclosures

Conflict of Interest: The authors declare no conflict of interest. All authors have read and agree with the contents of this paper.

Ethics Committee Approval: This work complies with the guidelines of the World Medical Association, Declaration of Helsinki. This work was approved by our institution's IRB as documented in the methods section

Strobe Statement: The authors have read the STROBE statement – checklist of items, and the manuscript was prepared according to its checklist of items.

Appendix: https://jag.journalagent.com/jilti/abs_files/JILTI-63935/JILTI-63935_(1)_JILTI-63935_(0)_Appendix_(son).pdf

Keywords

HCC; Network phenotyping strategy; PVT; survival

There has always been an awareness of the multiple component nature of the various influences on HCC growth (MTD) and invasion (PVT),[1–5] but we have hitherto not been able to integrate them well. Therefore, we used a Network Phenotyping Strategy or NPS,[6] which is designed to analytically process the integrated multi-component relationship facets of HCC seen in patient data, to guide us to a new way of understanding human HCC biology. NPS is derived to identify the stage of HCC disease at its actual "biological" duration during the index visit, at the point when the patient data were collected, by analyzing time-coherent relationship networks between personal clinical variable values. This novel non-statistical processing of extended clinical information in the patient data leads directly to consideration of HCC as a dynamic process, resulting in data-driven objective characterization of HCC stages and eventually, to better personalized outcome predictions.

To achieve this, a new principle of clinical data processing has been formulated. Instead of the standard statistical assumption, that HCC impacts patients more or less similarly, we analyzed HCC dynamics. NPS extracts information about the biological time of patient HCC, encoded in the relationship between coherently observed values of multiple variables. This allows the assigning of patients to objective progression-ordered clinical HCC stages, with earlier stages having good outcomes, whereas later clinical stages have poorer outcomes. This replaces the conventional outcome prognosis using an unknown future clinical state into gaining insight about the stage and prognosis from the known history of the patient's disease, encoded in the known personal biology at the patient index visit, as seen through the clinical variable relationship networks.

To achieve this, a new form of patient data, represented as K-partite graphs, was used. In these graphs, where relationships are not neglected, variable values are vertices, and disease progression time is captured by topology of value relationships, encoded by edges, connecting the patient's simultaneously observed data value vertices. Importantly, "prototype" data relationship network topologies are completely different in the earliest and latest HCC stages. That gives HCC dynamics monitored by the decrease of influence of early stage prototype-reflected biology and increase of influence of late stage prototype-reflected biology, both independently or in combination. The main discovery of this integrated approach to clinical data processing is that disease progression is not continuous. Instead, HCC patients emerged as naturally grouped into clinically well-separated groups with characteristic ratios of early and late stage biology contributions. We could then start applying the new objective causal stages of HCC to gain new insights into the role of interesting factors, relevant for diagnosis, treatment and prognosis for HCC.

Methods

Clinical Data

4802 HCC patients were identified from our previously published study[7, 8] who had a complete set of values for 17 standard baseline clinical parameters at initial clinical presentation, with tumor size and number and presence or absence of PVT, based on their initial CAT scan measurements and who also had known survival data. The 17 clinical parameters were chosen based upon baseline routine clinical data that is collected to evaluate any newly-presenting HCC patient and are in 3 groups: A), Demographics that included age, gender and HBV/HCV status; B), tumor characteristics, that included maximal diameter (MTD), tumor uni- or multi-focality, portal vein thrombosis (PVT) presence/absence and serum α-fetoprotein (AFP) levels; C), serum liver parameters and blood counts, including levels of albumin, total bilirubin, INR, ALT, AST, ALKP, GGT, Hb and platelets. This work for analysis of de-identified and diseased HCC patients was exempted from written informed consent. ITA.LI.CA database management conforms to Italian legislation on privacy and 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.

Mathematical Methods

NPS determines information about patient time of disease duration at index visit, τ_p from topology of 17-partite graph $\Gamma(\tau_p)$, representing all levels of the time-dependent relationships between all 17 personal coherently observed values (Fig. 1). In $\Gamma(\tau_p)$, each diagnostic variable is represented by a partition (ovals in Fig. 1). The partition contains the specific set of graph vertices, each representing consecutively ordered intervals of values, covering the entire physiological range of that variable. These intervals are defined by the heterogeneity in the histogram of the variable values, observed in the general HCC population.

The personal network of time-coherently observed data relationships is then represented by the network of edges (lines in Fig. 1), connecting the vertices in each partition, representing the observed values of 17 variables. The position of the connected vertex within the complete ordered set of vertices in a variable partition encodes the “first-order”, value-based, disease progression information, capturing if the observed value of one variable belongs to “early” or “later” stages of the disease. Edge connections of that variable value to remaining 16 observed variable values, represented by vertices in other partitions, captures the clinical context of the “earliness” or “lateness”, seen by the value of the single variable. The “shape/topology” of this network thus directly carries the personalized contextual information about τ_p, defining the disease progression stage.

NPS determines the τ_p from τ_x by computation, derived from 2 fundamental mathematical theorems, one showing that obtaining the best possible τ_p value requires maximization of Fisher information of the data,[9] and another showing that this is achieved by quantifying the differences in topologies between patient $\Gamma(\tau_p)$ and 17-partite prototype graphs.[10] By

specific method of finding these prototype topologies, NPS gains the most of new clinical content and explanatory transparency.

First, we examined whether 17-partite prototype networks are different in the earliest and the latest HCC stages. The clinical characteristics of these extreme HCC stages are properties of the disease biology. We therefore needed a network-based clinical disease descriptor. Its construction is straightforward, as we overlayed all individual patient partite graphs into one 17-partite multigraph D. It contains new clinical information through the multiplicities of edges between the variable values: they monitor the clinical relevance of respective observed relationships, quantified by high multiplicities of specific relationship edges in D. This reflects the functional preference for just some combinations of clinical value relationships over all other possibilities, caused by underlying HCC disease biology. We then formulated a "clinically ideal" feature of a networked marker of such HCC biology-determined topology prototype, since its identical copies will be found in D with high abundances. This clinical fact is translated into a "greedy" mathematical procedure by decomposing D into a minimal number of sub-multigraphs with identical multiplicities of all their edges.

An advantage of this approach is that such decomposition is unique[11, 12] for any given disease multigraph D, which can be understood that we obtain "principal topology subgraphs", potentially characterizing disease biology related markers in our data. For our 4802 patients, multigraph D is completely decomposed into 181 principal topology components, representing ~30 reduction in relationship dimensionality.

To find which of these 181 principal topology subgraphs characterize the earliest and the latest HCC stages, we applied the HCC specific adaptation of the general fact that outcome is significantly better for patients in earliest disease stages (HCC patients will be predominantly alive), compared to the outcome for patients in the latest HCC stages (patients will be predominantly dead, 63% in our cohort). We therefore looked for such principal topology prototypes in the set of 181, which, on one hand, share maximum relationship topology features with the most frequent topology features in the sub-group of surviving patients and simultaneously shared a minimal number of such relationships with the topological features, typical for deceased patient subcohort, and vice versa. This identified topology prototypes $P_m'(\tau_{min})$, P_m'' (τ_{min}) and P_x' (τ_{max}), P_x'' (τ_{max}).

We then used free energy information descriptors[13, 14] to independently quantify content X_p of P_x (τ_{max}) and content Y_p of P_y (τ_{min}) topology prototypes in a patient's observed networked data $\Gamma(\tau_p)$. This results in $[X_p,Y_p]$ NPS characterization of personal HCC stages in a 2-dimensional NPS map (Fig. 2).

For quantitative analysis and clinical interpretation of trends in the phenotype characteristics, we used the logarithm of abundance of P_y (τ_{min}) in the patient data, which we call (personal) disease burden, DB_p (τ_p). As the X_p and Y_p are expressed in (normalized) logarithmic scale, DB_p $(\tau_p)=Y_p-X_p$. From this definition follows that clinical burden has values between −1 and +1 and is negative for patients with data relationships consistent with dominance of early-stage markers and becomes positive for patients with

data relationships showing dominance of late-stage markers. The mathematical details are presented in Appendix 1.

Results

HCC Phenotyping by NPS

We found common networked data relationship "prototype" biomarker patterns P_y (τ_{min}), characteristic for the earliest stages of HCC, and other, completely different, common networked data relationship "prototype" patterns P_x (τ_{max}), characteristic for the latest stages. Different personal times from HCC disease onset are therefore observable by identifying the fractions of the early/late prototype relationship patterns in individual patient actual networked clinical baseline data. This permitted representation of HCC progression in a population as a 2-dimensional disease map, in which a patient is represented in the "HCC clinical progression plane" as a point, with coordinate Y_p, given by the personal content of early-stage HCC prototype topology (vertical axis) and by the content X_p of late-stage HCC prototype topology (horizontal axis), found in the patient's networked data.

We further showed that the distribution of patients in the NPS disease progression map is not uniform. Instead, patients with HCC are in this HCC clinical progression plane map tightly grouped into 25 well-separated sub-cohorts, within which patients are sharing most of their observed clinical feature relationships, with the frequency of their clinical differences decreasing exponentially from the most populated common HCC networked clinical profiles in the subgroup. This allowed us to identify 25 HCC NPS phenotypes sT,s=1…25, which are automatically ordered according to their characteristic disease progression stage (s=1 being the earliest, s=25 being the latest stage), shown in Figure 2.

We then examined the role of currently used clinical decision-supporting factors, in particular of PVT in the full context of other data in the NPS results, identifying and characterizing new clinically relevant categories of PVT patients. For this purpose, we used the fact that NPS analysis resulted in objective (non-statistical) partitioning of the clinically heterogeneous HCC population into stage-specific phenotype sub-populations, sT. These are naturally ordered according to the increasing time from disease onset (or, equivalently, by the increasing disease burden) and, at the same time, patients in sT's are "clinically normalized" to previously unachievable levels. We show evidence that sT's represent non-empirical, most likely disease biology-related stages of HCC.

We computed both phenotype-characteristic disease burdens $DB_s(\tau_s) = \frac{1}{N_s}\sum_{p=1}^{N} Y_p - X_p$ as the means of all personal disease burdens for patients from each sT phenotype, as well as the phenotype characteristic mean values $V_s(\tau_s) = \frac{1}{N_s}\sum_{p=1}^{N_p} V_p$ of any clinical variable. Plotting $V_s(\tau_s)$ against $DB_s(\tau_s)$ provided new clinical insights into HCC by revealing inter-dependencies between the overall disease burden and individual variables, as well as tools to validate that the sT's are indeed ordered by the characteristic increasing time from disease onset.

An example of such dependence, validating this order, is provided by plotting the phenotype-specific mortality as the function of specific disease burden, shown in Figure 3. The line fitted to this trend is a scaled Weibull cumulative distribution function[15] with a shape parameter equal to 2. This function describes the proportionality of deaths in the population to a power of time, providing mathematical confirmation of the theoretically expected functional interdependency.

Practical Example of using the NPS Web-Implementation

For an individual patient, values of all 17 variables recorded within the index visit are collected and entered into the Web tool. First, the diagnosis NPS model computes the compliances [X,Y] of patient's networked input with the early and latest stage biomarkers and shows it graphically by the patient's dot in the www. map. It then uses the result to determine the patient's sT phenotype. Assuming for example that the patient's computed [X,Y] falls into the 12T phenotype region of the map, the web tool shows the percent of deaths for patients in 12T in the training set. These mean mortality values are also shown in Table 1: A patient from 12T have ~65% chance of death.

For further personal mortality prognostication, the tool uses 2 pieces of additional clinical information. Firstly, we used follow-up information about time to death for subgroup of training cohort patients in each phenotype, who died within the follow-up interval to determine the mean time to death characteristic of each phenotype. Figure 5 shows that these mean times are inversely linearly related to the sT phenotype-characteristic mean disease burdens: patients who will die with the earliest stage HCC phenotypes have on average ~ 2 times longer time to death than patients in the latest stage HCC phenotypes. HCC stage specific times to death are summarized in Table 1. Thus, for our 12 T patient example, the mean time to death is about 34 months.

Secondly, using the www tool, we can further personalize the mortality prediction by using 25 independent NPS dead/alive optimized prognosticating models, one for each sT out of the 25. The input clinical data for 12T patient from our example are automatically re-submitted into the 12th NPS dead/alive prognostication model, resulting in newly computed {xD,yD} coordinates for the patient, which are again shown in second NPS map with only two regions. If the 12T patient NPS result {xD,yD} is in "alive" region, we can expect that he/she is in the smaller ~35% surviving subgroup in this phenotype. By consulting Table 1 for 12T, the reasonable survival time prognosis for this patient thus will be more than 34 months. For a patient with computed "dead" prognosis, 34 months will be typical time to death. Supplementary Figure 12 shows Kaplan-Meier curves for patients, classified as "alive" or "dead" by the HCC stage-specific NPS mortality prediction models.

PVT Phenotypes

Portal vein tumor thrombus in HCC is a manifestation of aggressive biology and poor prognosis (1–3) and is generally an exclusion criterion for several treatments, such as liver transplantation (4, 16) due to poor outcomes. It is also associated with poor liver function and high tumor recurrence rates after treatment. We therefore analyzed the group within this large cohort of non-surgical HCC patients who had PVT, searching for

possible useful prognostic subsets. First, we determined the characteristic PVT incidence $[PVT]_s = \frac{1}{N_s}\sum_{p=1}^{N_s} V_p(+)$ in every phenotype sT and plotted it against corresponding DB_S (Fig. 4a). It is clear that $[PVT]_s$ is a non-continuous function of disease burden. The 25 primary sT phenotypes form 5 broader groups, shown by differently colored boxes, each with well-separated range of practically progression-independent $[PVT]_s$ percentages.

The explanatory insight into this different behavior of a major negative prognostic factor is provided by projecting the constituent basic phenotypes from every PVT group, shown by colored arrows in Figure 4b. The content of the P_x (τ_{max}) in each patient profile is shown on the horizontal (X) axis. The HCC progression in the respective PVT phenotypes proceeds vertically, top-down, in the respective columns of the sT fundamental phenotypes (The fluctuations in the last 2% PVT phenotypes are explainable as “noise”).

The meaning of vertical top-down progression 1T→2T→4T→6T→8T→11T of the fundamental phenotypes sT in the green PVT group is that, on average, those patients have a small constant content of the “bad” P_x (τ_{max}) relationship markers (x-coordinate is not changing in the progression). This shows that gradual disappearance of the “good” P_y (τ_{min}) relationship features as HCC goes from top (early) to down (later) stages exclusively monitors the HCC progression in this sequence of sT’s.

Thus, the patients from green PVT cohort present with minimal and constant evidence of the “bad” HCC biology, related to the latest HCC stages. The patients from blue PVT subgroup present with a larger constant content of the “bad” P_x (τ_{max}) than the green PVT subgroup. Thus, the biological heterogeneity of blue PVT patient subgroup is higher than that of the green PVT patient subgroup. The yellow PVT patient subgroup contains even larger constant evidence of their “bad” biology than the blue PVT subgroup (and of course much higher than green PVT subgroup). The same holds (with some noise, which is relatable to the dominancy of P_x (τ_{max}) and minimal content of P_y (τ_{min}) for brown and red PVT subgroups.

In summary, each PVT subgroup has a constant, characteristic increase in content of P_x (τ_{max}), which defines the constant frequencies of occurrence of PVT in them. The P_x (τ_{max}) contains the relationships between values of 17 variables, which are characteristic for late stages of HCC. This is responsible for the increase of PVT pincidence in the sequence of the 5 PVT groups, progressing from green to red. This is an example of an important function of NPS, because its results are not statistical, but derived from functional clinical first principles, they are hypothesis-generating, directly providing information for designing validation trials. For using this result in our practical example, we identify the patient 12T phenotype in Figure 4a, assigning him to “brown” group with 18–20% incidence of PVT.

Discussion

The novelty of NPS methodology in clinical analytics originates from integrating all the patient- and time-coherently observed relationships between multiple clinical variable values into K-partite graph-based processed input. Simultaneously, new clinical content in NPS results becomes available since we integrated the appropriate tools of discrete mathematics,

general physics and theoretical informatics to obtain the implementable and completely scalable functional, non-statistical mathematical processing of this “new” information in “old clinical data”, without encountering the “complexity catastrophe” that would be encountered by conventional methodologies.

Another novel feature of NPS approach is change of its founding clinical principle from conventional frequentist’s paradigm of average impact of disease on any patient to determining the personal time of disease duration τ_p at the index clinical visit when the data are collected. This enables explicit characterization of disease as a dynamic process using apparently “static” clinical data. Here we show by characterizing the HCC progression that networked relationships between coherently observed values of multiple informative clinical variables encode enough information about the personal disease progression history for NPS to extract τ_p from their topology.

This is mainly due to the existence of networked data relationship “prototype” biomarker patterns P_y (τ_{min}), characteristic for the earliest HCC stages, and completely different, common networked data relationship “prototype” biomarker patterns P_x (τ_{max}), characteristic for the latest HCC stages. Different personal times τ_p from HCC disease onset are therefore observable by identifying the contributions of the early/late topology prototype relationship biomarker patterns to the patient’s actual networked clinical baseline data. This result allows the representation of the HCC dynamics as a 2-dimensional disease progression map, in which a patient is represented as a point [X_p, Y_p], with coordinates, given by the personal content of early-stage HCC prototype topology (vertical axis) and content of late-stage HCC prototype topology (horizontal axis), found in the patient’s integrated and networked data.

The results of these NPS data processing steps represent discoveries of new clinical features, providing new insights into HCC biology or allowing formulation of new testable hypotheses about HCC biology. We identified 25 objectively defined HCC phenotypes sT, each with clearly distinguishable clinical characteristics, which determine the τ_s-defined order of these stages (from s=1 to s=25). The main discovery of NPS is that the disease burden, imposed upon the patient by HCC stage sT, is not continuous, but proceeds by stepwise transitions between “clinically stable” disease stages.

Applicability of this NPS result in diagnostic clinical practice stems from the fact, that there are significantly large sub-populations of HCC patients in each sT HCC stage phenotype.

Applicability of this NPS result in prognostic clinical practice stems from disease progression-defined ordering of the NPS phenotypes. If we determine that patient is in HCC stage sT at baseline, then the NPS phenotype ordering indicates that the disease progression for that specific patient will be either stable disease or progression to (s+1)T stage or (possibly treatment-related) improvement to (s-1)T.

This set of τ_s-ordered HCC phenotype subsets also provide detailed insight into HCC biology by allowing us to examine changes in the observable functional background and the presence of established risk and treatment-decision factors as a function of disease progression. We used the example of presence of PVT, a major negative prognostic factor

for HCC. Therefore, clinical predictive factors for presence of PVT and identification of patient subgroups having PVT with differing survivals is a crucial part of HCC patient management. We hypothesize that this might be biologically related to the reported stem cell marker heterogeneity, which is open to experimental validation. So far, the main identified HCC patient risk factors for PVT development include large (>5cm) HCC size, elevated AFP levels (>500 IU/mL), low serum albumin levels and male gender, while predominant prognostic factors include serum AFP and albumin levels and indices of inflammation.[17–19]

Our NPS may also help explain different survival outcomes for patients having the same treatment. This may be a reflection of known tumor heterogeneity and evolution that has derived from multiple studies of histopathology, tumor mutation burden and satellite instability. We know that tumors are not static in their composition. Our findings reflect that. There are likely multiple pathways to tumor growth, as well as changes in the pathways during various phases of growth of a tumor in an individual patient. Our τ_s-ordered stages reflect quantitatively the clinical observations of these events. It is also true than in most series, the percent of patients with PVT increases with MTD up to about 50%, but there are still about 30% of patients with large MTD >8cm without PVT.[20] Perhaps, the same stem cell that causes increasing MTD might also (with input from other factors) cause an increase in PVT. Or perhaps a patient's HCC has different stem cells.

Conclusion

NPS uses a personal networked relationship-based characterization of the patient clinical status, which is never partitioned (step 1). From these personal networks, NPS constructs the characterization of HCC disease, allowing a determination of prototype-networked biomarkers for earliest and latest HCC stages (step 2). These 2 steps result in the possibility of characterizing the personal disease progression stage at baseline, by determining compliance of patient actual clinical data network with early-stage prototype networked biomarker (compliance descriptor Y) and with final stage prototype networked biomarker (compliance descriptor X). Patient HCC progression is then defined by the [X, Y] stage point in the 2D-HCC progression map (step 3). This map directly shows that HCC disease appears to be stable in 25 clinically relevantly large sub-populations of patients 1T… 25T. The stable HCC disease of a patient in stage sT progresses to different stage rT only after the accumulation of a critical amount of new personal relationships, which will change the 2 personal compliances [X, Y] of the personal networked profile to those in another phenotype rT, which is different from sT. Further improvement of clinical insight into HCC biology can be obtained in future by integrating additional coherently acquired patient data with our 17 by the NPS approach, which can seamlessly integrate new types of clinical data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding:

This work was supported in part by NIH grant CA 82723 (B.I.C).

Abbreviations:

NPS	Network Phenotyping Strategy
MTD	Maximum tumor diameter
PVT	Portal vein thrombosis by tumor
τ_p	Time of disease duration at index visit
$\Gamma(\tau_p)$	Personal 17-partite graph
HBV/HCV	Hepatitis
AFP	serum. α-fetoprotein
INR	international normalized ratio
ALT	alanine transaminase
AST	aspartate transaminase
ALKP	alkaline phosphatase
GGT	γ-glutamyl transferase
Hb	hemoglobin

References

1. Nakashima T, Okuda K, Kojiro M, Jimi A, Yamaguchi R, Sakamoto K, et al. Pathology of hepatocellular carcinoma in Japan: 232 consecutive cases autopsied in ten years. Cancer 1983;51:863–77. [PubMed: 6295617]
2. Mahringer-Kunz A, Steinle V, Kloeckner R, SchoIen S, Hahn F et al. The impact of portal vein tumor thrombosis on survival in patients with hepatocellular carcinoma: A cohort study. PLoS One 2021; 16: e0249426. [PubMed: 33961627]
3. Chen K-L, Gao J Factors influencing short-term and long-term survival of hepatocellular carcinoma patients with portal vein tumor thrombosis who underwent chemoembolization. World J Gastroenterol 2021;27:1330–1340. [PubMed: 33833486]
4. Minagawa M, Makuuchi M. Treatment of hepatocellular carcinoma accompanied by portal vein tumor thrombus. World J Gastroenterol 2006;12:7561–7567. [PubMed: 17171782]
5. Akkiz H, Carr BI, Kuran S, Karaogullarindan U, Uskudar O et al. Macroscopic portal vein thrombosis in HCC patients. Can J Gastroenterol Hepatol 2018;3120185. [PubMed: 30009156]
6. 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(2):339–46. [PubMed: 25843738]
7. Carr BI, Pancoska P, Giannini EG, Farinati F, Ciccarese F, Rapaccini GL, Marco MD, Benvegnù L, Zoli M, Borzio F, Caturelli E, Chiaramonte M, Trevisani F; Italian Liver Cancer (ITA.LI.CA) group. Identification of two clinical hepatocellular carcinoma patient phenotypes from results of standard screening parameters. Semin Oncol 2014;41(3):406–414. [PubMed: 25023357]
8. Carr BI, Guerra V, Donghia R, Farinati F, Giannini EG, Piscaglia F, et al. Changes in hepatocellular carcinoma aggressiveness characteristics with an increase in tumor diameter. Int J Biol Markers 2021;36(1):54–61. [PubMed: 33641486]
9. Wu JQ, Horeweg N, de Bruyn M et al. Automated causal inference in application to randomized controlled clinical trials. Nat Mach Intell 4, 436–444 (2022).

10. Frieden BR, Gatenby RA. Principle of maximum Fisher information from Hardy's axioms applied to statistical systems. Phys Rev E Stat Nonlin Soft Matter Phys 2013;88(4):042144. [PubMed: 24229152]
11. Banks D, Constantine G Metric Models for Random Graphs. J. of Classification 15, 199–223 (1998).
12. Messuti S, Rödl V, Schacht M: Packing minor-closed families of graphs into complete graphs, Journal of Combinatorial Theory, Series B, Volume 119, 2016, (245–265), ISSN 0095–8956, 10.1016/j.jctb.2016.03.003.
13. Friston K The free-energy principle: a unified brain theory? Nat Rev Neurosci 11, 127–138 (2010). 10.1038/nrn2787 [PubMed: 20068583]
14. Gottwald S, Braun DA (2020) The two kinds of free energy and the Bayesian revolution. PLoS Comput Biol 16(12): e1008420. 10.1371/journal.pcbi.1008420 [PubMed: 33270644]
15. Li X, Marcus D, Russell J, Aboagye EO, Ellis LB, Sheeka A, Park WE, Bharwani N, Ghaem-Maghami S, Rockall AG. Weibull parametric model for survival analysis in women with endometrial cancer using clinical and T2-weighted MRI radiomic features. BMC Med Res Methodol 2024;24(1):107. [PubMed: 38724889]
16. Aydin C, Yilmaz S. Is Macroscopic Portal Vein Tumor Thrombosis of HCC Really an Exclusion for Liver Transplantation? J Gastrointest Cancer 2020;51(4):1137–1140. [PubMed: 32833219]
17. Carr BI, Guerra V, Donghia R, Yilmaz S. Tumor multifocality and serum albumin levels can identify groups of patients with hepatocellular carcinoma and portal vein thrombosis having distinct survival outcomes. Ann Med Surg (Lond). 2021; 66:102458. [PubMed: 34141428]
18. Carr BI, Guerra V, Donghia R. Portal Vein Thrombosis and Markers of Inflammation in Hepatocellular Carcinoma. J Gastrointest Cancer 2020;51:1141–1147. [PubMed: 32851544]
19. Carr BI, Guerra V. Low Alpha-Fetoprotein Levels Are Associated with Improved Survival in Hepatocellular Carcinoma Patients with Portal Vein Thrombosis. Dig Dis Sci 2016;61:937–47. [PubMed: 26576554]
20. Carr BI, Bag HG, Akkiz H, Karaoğullarindan Ü, Ince V, Isik B, Yilmaz S. Identification of 2 large size HCC phenotypes, with and without associated inflammation. Clin Pract (Lond) 2022;19(4):1953–1958. [PubMed: 37621527]

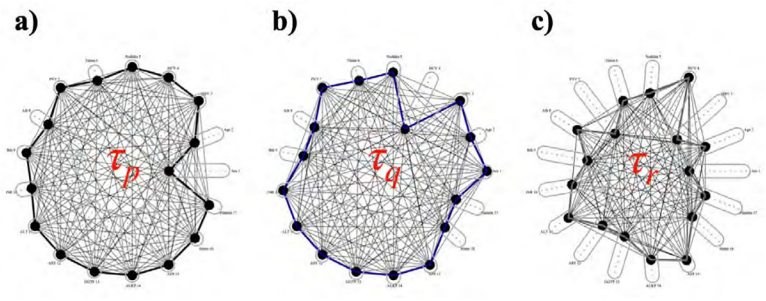

Figure 1.

17-partite graphs $\Gamma(\tau_x)$ for three patients (p, q and r) at different stages of HCC. Ovals - individual clinical variables. Circles in ovals - vertices, representing variable values or intervals of each invariable. Vertices are naturally ordered: values, typically observed at earlier disease stages are on the outside and values observed typically at late stages are on the inside ends of each partition. Solid circles - vertices, representing observed values of respective variables for the patient. Lines - network of relationships between coherently observed values of all variables. Figure demonstrates the change of patient's $\Gamma(\tau_x)$ topology when HCC progresses from early (patient p) through intermediate (patient q) to late (patient r) stages.

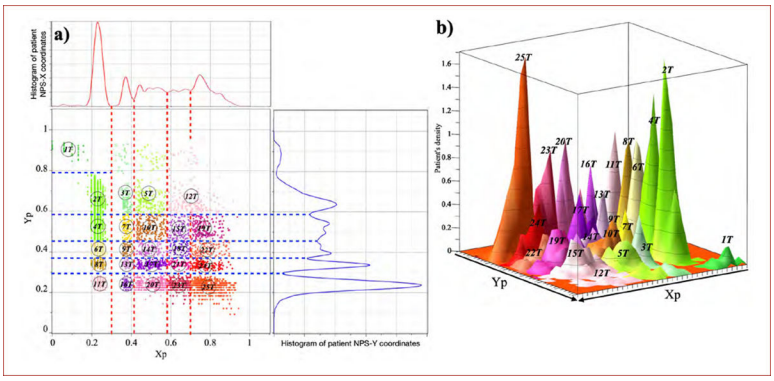

Figure 2.

(a) Central panel: NPS 2-dimensional map of disease progression (see text for explanation of clinical meaning of horizontal and vertical axes). Each patient out of 4802 is represented by a point [X_p, Y_p]. Right panel - histogram of patient density in the top-down direction of decreasing contribution of early HCC stage biology. Top panel - patient density in the direction of left-right direction of increasing contribution of latest HCC stage biology. Dotted lines - boundaries of HCC progression, separating patient sub-populations in respective NPS phenotype stages sT. **(b)** 3-dimensional histogram of patient population distribution in all 25 HCC NPS phenotype stages sT.

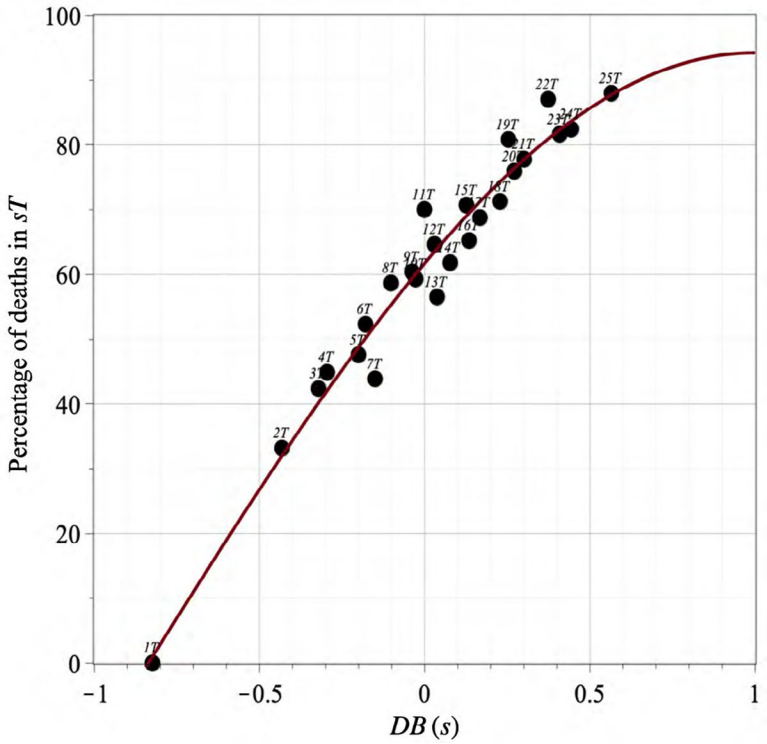

Figure 3.

Relationship between the mean disease burden DB[s] and mean mortality for all 25 HCC NPS phenotype sub-populations sT,(s=1…,25)(circles). Least squares fit of the relationship by the Weibull function (lines).

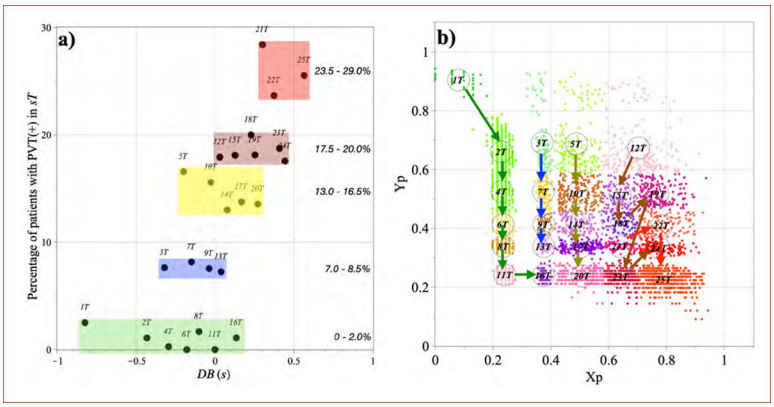

Figure 4.

(a) Relationship between the mean disease burden DB[s] and mean PVT incidence for all 25 HCC NPS phenotype sub-populations sT,(s=1…,25). Boxes - patients with specific NPS HCC phenotype stages, constituting the five PVT phenotypes. **(b)** Projection of HCC progression in respective PVT phenotypes. Arrows are colored as the boxes in Fig. 4a.

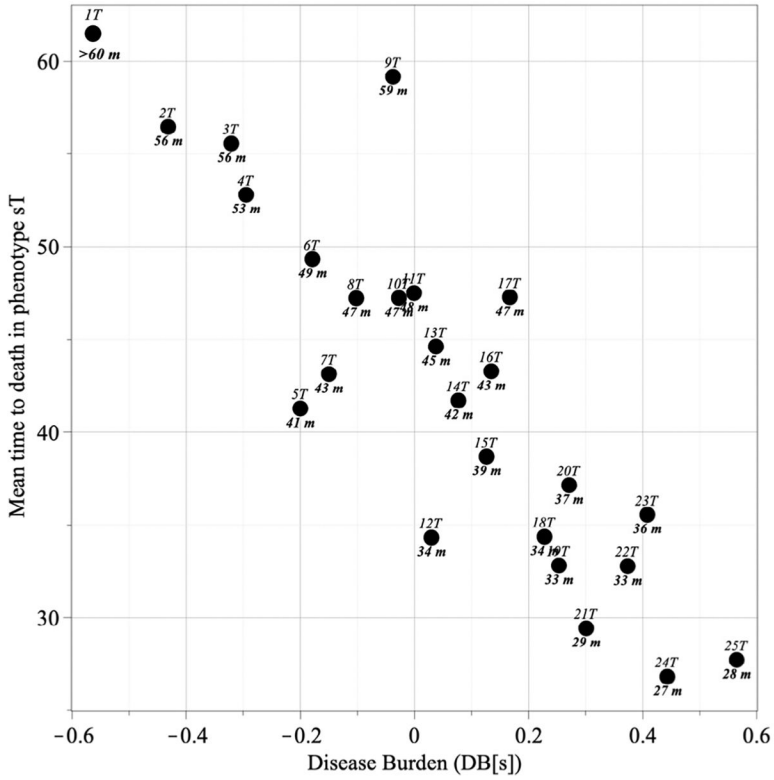

Figure 5.

Relationship between the mean disease burden DB[s] and mean time to death (in months) for all 25 HCC NPS phenotype sub-populations sT,(s=1…25). sT = labels above the results, mean time to death = labels below the results. See also Table 1.

Table 1.

A patient from 12T have ~65% chance of death

	1T	2T	3T	4T	5T	6T	7T	8T	9T	10T	11T	12T	13T	14T	15T	16T	17T	18T	19T	20T	21T	22T	23T	24T	25T
Mean TtD [months]	>60	56	56	53	41	49	43	47	59	47	48	34	45	42	39	43	47	34	33	37	29	33	36	27	28
Mortality [%]	0.0	33.2	42.4	44.9	47.6	52.3	43.9	58.7	60.4	59.3	70.0	64.6	56.5	61.8	70.7	65.2	68.8	71.2	80.8	75.9	77.8	87.0	81.6	82.4	88