



<http://www.diva-portal.org>

This is the published version of a paper published in *Proceedings of the National Academy of Sciences of the United States of America*.

Citation for the original published paper (version of record):

Bidkhor, G., Benfeitas, R., Klevstig, M., Zhang, C., Nielsen, J. et al. (2018)
Metabolic network-based stratification of hepatocellular carcinoma reveals three
distinct tumor subtypes
Proceedings of the National Academy of Sciences of the United States of America

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-248680>

Biological Sciences, Systems Biology

Metabolic network stratification of hepatocellular carcinoma reveals three distinct tumor subtypes

Running title: Network-based stratification of HCC

Gholamreza Bidkhor^{1,4}, Rui Benfeitas^{1,4}, Martina Klevst², Jens Nielsen³, Mathias Uhlen¹, Jan Boren², Adil Mardinoglu^{1,3}

¹ Science for Life Laboratory, KTH - Royal Institute of Technology, SE-171 21, Stockholm, 8 Sweden

² Department of Molecular and Clinical Medicine, University of Gothenburg, and 12 Sahlgrenska University Hospital, Gothenburg, Sweden

³ Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

⁴ These authors contributed equally.

Gholamreza Bidkhor (gholamreza.bidkhor@scilifelab.se); Rui Benfeitas (rui.benfeitas@scilifelab.se); Martina Klevst (martina.klevst@gu.se); Jens Nielsen (nielsenj@chalmers.se); Mathias Uhlen (mathias.uhlen@scilifelab.se); Jan Boren (Jan.Boren@wlab.gu.se); Adil Mardinoglu (adilm@chalmers.se)

Correspondence:

Adil Mardinoglu

Science for Life Laboratory,
Box 1031, Solna,
SE-17121, Sweden
adilm@chalmers.se

Keywords: hepatocellular carcinoma, biological networks, systems biology, personalized medicine, genome-scale metabolic model

SUMMARY

Hepatocellular carcinoma (HCC) is one of the most frequent forms of cancer and effective treatment methods are limited, with challenges related to its large heterogeneity. A great need exists for comprehensive approaches to stratify HCC using methods capable of incorporating inter-tumor variability, while providing biologic insights and ultimately identifying suitable therapeutic targets in an individualized manner. Here, we have employed a novel metabolic network-based stratification of HCC which uses modeling and network topology/controllability to stratify and characterize hundreds of samples based on transcriptomic data. The comprehensive analysis identified three distinct HCC subtypes with substantial metabolic differences, extending also to distinct genomic, gene expression, and immunohistochemical differences. These subtypes show large differences in clinical survival, associated with altered Kynurenine metabolism, WNT/ β -catenin-associated lipid metabolism alterations, and PI3K/AKT/mTOR signaling. The gene expression analysis show that the three groups rely on alternative enzyme-coding genes (e.g. ACSS1/ACSS2/ACSS3, PKM/PKLR, ALDOB/ALDOA, MTHFD1L/MTHFD2/MTHFD1) to drive the same reactions. We have also identified 8 – 28 subtype-specific genes with pivotal roles in controlling network and whose *in silico* silencing shows that these could be potential new drug targets for one of the iHCCs. Finally, we have experimentally observed opposite expression patterns between genes expressed in high/moderate and low survival tumor groups in response to hypoxia, reflecting promoted hypoxic behavior in patients with poor survival. Overall, our analyses show that the substantial HCC heterogeneity can be stratified using a metabolic-network driven approach and this stratification can have clinical implications as it can drive the development of personalized medicine.

SIGNIFICANCE

Hepatocellular carcinoma (HCC) is a highly heterogeneous and deadly form of liver cancer. Here, we characterized and stratified HCC tumors based on genome-scale metabolic network heterogeneity. Our newly developed *in silico* method enabled the identification of three HCC subgroups with distinct metabolic, signaling and survival properties, as well as hypoxic-driven gene expression responses. We verified the results of our analysis by performing additional experiments and associated it with patient survival. We further identified a number of subgroup-specific genes pivotal in controlling the entire metabolism and discovered genes that can be targeted for development of efficient treatment strategies for specific patient group. Our systems level analyses provided a systematic way for characterization of liver cancer sub-types.

INTRODUCTION

Hepatocellular carcinoma (HCC) is a prevalent form of liver cancer, the third leading cause of cancer-related worldwide mortality, and its incidence is predicted to increase globally (1). However, due to this disease's large heterogeneity, a complete understanding of the biological phenomena underlying HCC onset and progression remains elusive. Comprehensive approaches

capable of incorporating inter-tumor variability, while providing biologic function insights are thus of great need for understanding biological phenomena and identifying suitable therapeutic targets.

Systems biology approaches have been pivotal in tackling this challenge in cancer. Genomic, transcriptomic, or metabolic characterizations of HCC consisting of large-scale data are currently available (2-8). This enabled the identification of markers associated with recurrence and poor prognosis (9-11). In turn, Genome-scale Metabolic Models (GEMs), comprehensive metabolic network descriptions incorporating reaction stoichiometry information and functional descriptions, have been successfully used to metabolically characterize HCC, as well as identify targets for personalized treatment (7, 12-14). For instance, HCC tumors display altered acetate metabolism responses in patients with differential patient survival (7). Analysis of HCC metabolism also pointed to potential anticancer metabolite analogues that would not be toxic for noncancerous liver tissues (12), and to substantial association and antagonistic responses between redox and central metabolisms (14). These observations indicate the clear strengths in integration of large-scale omics data with personalized medicine approaches. However, while these methods implicitly consider metabolic network structure, they do not permit stratifying tumors based on network heterogeneity itself, and instead rely on identification of key genes/metabolites and tumor stratification based on their levels. In turn, topology-driven network analyses, including protein-protein interaction, signaling, gene regulatory and metabolic networks (15-17) provide an alternative view over cancer networks. For instance, network analysis has identified essential proteins from a lethality perspective, as well as those capable and indispensable for controlling network (18-22). However, topology-driven methods do not take into account biological functionality, one important strength in GEM-driven and similar analyses.

Here, we integrate multi-omic data with modeling and metabolic network-based analysis to introduce a whole network-driven stratification of HCC tumors. Consistent tumor stratification was performed across different datasets consisting of hundreds of HCC tumors. Importantly, though this analysis considers only metabolic network information, substantial differences are observed at the gene expression, genomic, clinical, and survival level. Additionally, we identify novel HCC subtype-specific therapeutic targets that have important roles in controlling cancer network not noncancerous liver samples. Finally, we have experimentally observed that expression of genes associated with good and poor prognosis tumors shows opposite responses to hypoxia.

RESULTS

Characterizing metabolic heterogeneity and identifying controlling genes in HCC

We started by retrieving the transcriptomic and clinical data for 369 HCC individuals, along with 50 non-cancer liver samples from GDC (8). This dataset was split in 2 parts: a test set, consisting of 186 patients with detailed clinical information for clinical and signature data analysis, and a set consisting of 183 patients that was used later for validation. We integrated the test set with an HCC-specific genome scale metabolic model (12) to generate patient-specific HCC and non-tumor

GEMs (see Methods). After excluding non-functional models, we constructed personalized directed *functional* gene-gene networks (fGGNs), a novel approach introduced here for clarifying metabolic gene importance in HCC (Fig. S1) inspired by a previous approach (23). In fGGNs, nodes represent metabolic genes (enzymes) and edges represent connections between metabolic genes that are formed if a metabolite product of a gene's reaction serves as substrate for the reaction driven by the other gene (Fig. 1A).

After validation of fGGNs against randomly generated networks (Fig. S2A), we sought to compare heterogeneity across patients by testing model similarity within and between HCC vs non-cancer fGGNs. We investigated nodes based on their control over the network, shortest connections with other nodes, number of neighboring nodes, and direction of interaction by computing the centrality topology parameters betweenness, normalized closeness, eccentricity and degree (Table S1). Comparison of these scores within HCC and non-cancer group indicates that the former group is substantially more heterogeneous than the latter, where the median node absolute deviation for each of the parameters tends to be larger in HCC than non-tumor samples (Fig. 1B and C). In turn, between-group comparison shows substantial differences between HCC and non-tumor samples expressed at the network level (Fig. S2B). Overall, all tested parameters show that non-cancer fGGN are more similar to each other in comparison with HCC networks at the network level.

We then aimed to identify genes that are pivotal in controlling of the full networks through network controllability approaches (i.e. minimum driver node sets, MDS). Most of the 224 MDS genes identified are involved in transport reactions, fatty acid metabolism, oxidative phosphorylation, nucleotide metabolism and carnitine shuttle (Table S2). Similarly to previous approach (18), we classified nodes based on their controllability classification as indispensable, neutral and dispensable, i.e. those whose removal from the network respectively increase, do not change, or decrease the minimum number of MDS (Table S2). We identify 188 genes that are indispensable in $\geq 80\%$ of the fGGNs. Our observations indicate that indispensable genes show very high degrees indicative of high connectivity in both HCC and noncancerous networks (Fig. 1D). For instance, indispensable, neutral and dispensable genes show median degrees of 32, 17 and 9 in HCC, respectively, whereas they show median degrees of 34, 22 and 11 in noncancerous networks. Importantly, dispensable and indispensable nodes may show similar degrees indicating that not all highly connected genes (i.e. hubs) are controlling the network, but most indispensable genes are hubs.

In silico gene silencing of all 2892 metabolic genes in GEMs shows that $>95\%$ of HCC samples show no growth when MDS or indispensable genes are silenced, much higher than the observed fractions for silencing of other genes ($<50\%$) (Fig. 1E). Together, MDS and indispensable genes are hereafter called “controlling genes” based on their role in network control. Based on the controllability and MDS classifications we observe clear separation of HCC and non-cancer fGGNs as indicated by Principal Component Analysis (Fig. 1F), otherwise not achieved when solely considering gene expression (Fig. S2C). These observations show that despite the high heterogeneity expressed at the gene level in HCC, network analyses identify distinct and important

genes that may be used to efficiently separate HCC and non-cancer samples based on network controllability.

Network-based stratification reveals metabolic, survival and cancer hallmark differences in HCC

Having identified important differences between HCC and noncancerous fGGNs, we then used a novel network-based approach to stratify the patients. Here, we introduce the utilization of *functional* gene-gene networks to stratify tumors based on gene expression data, using techniques previously employed to stratify tumors based on somatic mutations (24). We combined the personalized fGGN into a single generic fGGN representative of the features of all 186 patients, consisting of 1972 metabolic genes (see Methods), that was used for stratification. Integrating patient transcriptomic data with the generic fGGN, and employing network smoothing to spread the influence of each expression profile on the neighborhood of the network, we generated expression profiles that reflect the fGGN structure. These expression profiles were subsequently stratified using Nonnegative Matrix Factorization (NMF). An optimum number of three HCC groups was identified (Fig. S2D), each consisting of 85, 49 and 52 patients each (iHCC1 – 3) with substantial gene expression, biological process and clinical differences (Fig. 2, Table S3).

Differential expression analysis (Fig. 2A) identifies 2409 differentially expressed genes between iHCC2 vs iHCC3, 2318 genes between iHCC1 vs iHCC3, and 1115 genes between iHCC1 vs iHCC2 ($Q < 0.05$, DESeq, Table S3). Cancer hallmark gene set enrichment analysis (25) highlights significant differences in hallmarks of cancer ($Q < 0.01$, Fig. 2A). For instance, iHCC3 displays upregulated E2F targets, mTOR, MYC, inflammatory response, mitosis, G2M checkpoint, and DNA repair in when compared with iHCC1/iHCC2. In turn, iHCC2 shows WNT/beta catenin activation. Mitosis and cell cycle-associated gene expression is downregulated in iHCC2 in comparison with iHCC1/iHCC3 and inflammation is higher in iHCC1/iHCC3.

Among the genes differentially expressed between the low and high survival iHCC3 and iHCC1 groups, we identified several prognostic markers (DESeq, Fig. 2B, Table S3). For instance, when compared with iHCC3, tumors from iHCC1 display upregulated expression of 64 favorable prognostic markers and downregulated expression of 45 unfavorable prognostic markers (Fig. 2B, $Q < 0.05$, DESeq), among the 469 metabolic genes previously identified as prognostic markers in liver cancer (11). In turn, iHCC2 shows mixed up- and downregulation of these prognostic markers. iHCC3 tumors additionally present downregulated expression of 123 (out of 157) liver-specific genes ($Q < 0.05$, Table S3), upregulation of genes associated with immune signatures (26) and genes associated with metastasization such as HIF1 α , IL1, TNF α , NF κ B, and TGF β are upregulated in iHCC3 (Table S3). Survival differences of the 3 groups are consistent with expression of prognostic markers, where iHCC1 presents the highest survival rate, followed by iHCC2, and iHCC3 (Log rank $P < 0.001$, Fig. 2C).

Though differences are observed between the three groups, iHCC3 tumors are markedly distinct from those of iHCC2 and iHCC1. A larger number of genes are differentially expressed between iHCC3 vs iHCC1/iHCC2 when compared with iHCC1 vs iHCC2, and several cancer hallmarks are simultaneously enriched in iHCC3 in comparison with either iHCC1 or iHCC2 (Fig. 2A). For instance, iHCC3 tumors present downregulated oxidative phosphorylation, fatty acid metabolism, adipogenesis, and upregulated DNA repair, G2M checkpoint, epithelial to mesenchymal transition and inflammation, when compared with iHCC1 or iHCC2 (Table S3).

Gene set enrichment analysis performed in PIANO (25) using biological processes retrieved from MSigDB highlights iHCC-specific responses (Table S4). For instance, iHCC1 displays upregulated tryptophan and indole metabolism, but downregulated ncRNA metabolism, and ribosome biogenesis ($Q < 0.05$), when compared with tumors of iHCC2 and iHCC3. Tumors in iHCC2 displays ($Q < 0.05$) upregulated heme, glutamine metabolism, drug metabolism and transport, and oxidative demethylation, but downregulated cell development and GPCR signaling, when compared with iHCC3 and iHCC1. Tumors in iHCC3 show the largest changes in biological processes when compared with iHCC1 or iHCC2, with upregulation of multiple processes associated with cell proliferation, cell cycle progression and mitosis, development, chromosome segregation, cytoskeleton organization, immune response, DNA replication and recombination ($Q < 0.05$). In turn, iHCC3 displays downregulated fatty acid β oxidation, lipid oxidation, small molecule and catabolism, and metabolism of several amino acids including glycine, glutamate, glutamine, serine, aspartate, drug catabolism and response to xenobiotic stimulus (Table S4). Consistent with the substantial differences between iHCC3 and the two other tumor groups, iHCC2 tumors show similar metabolic behavior to those of iHCC1 (Table S5), and their gene expression is more similar to those of iHCC1 than to those of iHCC3 (Fig. 2D, mean Spearman's $\rho \approx 0.9$ iHCC1 vs iHCC2, < 0.8 iHCC3 vs iHCC1 or iHCC2).

Importantly, our stratification method highlights several *stratifying* genes whose expression is substantially different between the 3 iHCC groups. This is the case of XDH, KMO, TDO2 and SC5D in iHCC1; GLUL, AQP9, RHBG, SLC1A2, SLC13A3, ACSS3, AOX1 and CYP3A4 in iHCC2; and PKM, G6PD, PGD, ENO1, SRM, and ALDOA in iHCC3 (Fig. 3A and B, Fig. S3). Other genes such as MTHFD1, ALDH6A1, and ACSM2B are similar in both iHCC1 and iHCC2, but differ significantly in comparison with iHCC3.

Revealing the association between metabolism, recurrence signatures, Wnt/ β -catenin and PI3K/Akt/mTOR signaling

The above results show that the iHCC subgroups present specific features at the survival, gene expression, prognostic marker, and metabolic level identified solely based on analysis of metabolic gene networks. These tumors are also differentially associated with known HCC properties such as HIPPO signature, hypermethylation, DNA copy number, cholangiocarcinoma-like traits (2), RS65 gene-based risk scores (27), and HB16 signature (3) (Fig. 3A, Table S3). For instance, 84% of iHCC2 subjects are men (vs ~50% in other iHCCs), and about half of the patients in iHCC2 and

iHCC3 display alcoholic liver disease, much higher than the <25% observed in iHCC1 ($Q < 0.01$). Additionally, iHCC2 tumors also show lower genome doubling, higher hypermethylation and CDKN2 silencing (Fig. 3A, $Q < 10^{-4}$), and all iHCC2 tumors show AFP < 300 ng/mL. iHCC1 and iHCC2 tumors are associated ($Q < 10^{-4}$, Chi-square test) with markers of hepatocyte differentiation (>54% tumors display Hoshida 3)(4), and maturity (>79% HB16 C1). In turn, no iHCC3 tumors show differentiation markers (0% Hoshida 3) and instead are associated with known markers of low survival ($Q < 0.05$, Chi-square test, Fig. 3A, Table S3) including NCIP score A (>96%), high recurrence risk SNUR (>76%) (10), and high expression of recurrence risk marker CD24 (Log fold change ≈ 2.55 for comparison vs iHCC1, $Q < 0.00085$). The lower survival and predominance of aggressive tumors in iHCC3 may be associated with the significantly ($Q < 0.02$, Chi-squared test) larger proportion of advanced tumors in this group (>51% Grade 3, <49% Grades 1 and 2) compared with iHCC2 (30% G3 and <70% G1 and G2) or iHCC1 (<22% G3, >77% G1 and G2). iHCC2 also shows altered cytochrome P450 and xenobiotic metabolism in comparison with the 2 other clusters (Fig. S9, Fig. S10).

Interestingly, several observations associate altered Wnt/ β -catenin, PI3K/Akt/mTOR signaling, with the novel iHCC phenotypes described here. Most iHCC3 tumors are associated with MYC and AKT activation as indicated by the high incidence of Hoshida 2 (in 96% of tumors, Fig. 3A). Additionally, we identified (28) the top-25 genes co-expressed with stratifying/controlling genes in each iHCC for 360 TCGA tumors, and observe positive co-expression of AKT1 and MTOR and stratifying/controlling genes in iHCC3 and their co-expressed genes (Pearson's $r > 0.32$, $Q < 0.01$, Fig. 4). AKT1 and MTOR are negatively co-expressed with stratifying/controlling genes in iHCC1 and iHCC2. In turn, Hoshida signatures are not substantially different between iHCC1 and iHCC2 (22% and 11% Hoshida 1 respectively, $Q > 0.3$, Chi-square test). However, the 5 following observations suggest a strong association between disturbed Wnt signaling and the iHCC2 phenotype. First, 75% iHCC2 tumors show mutations in CTNNB1, a gene that codes for β -catenin in the Wnt pathway (Fig. 3A), substantially higher than <13% observed in iHCC1 and iHCC3 ($Q < 10^{-5}$, Chi-square test). Second, iHCC2 tumors also show upregulated expression of β -catenin target genes, for instance glutamine synthetase GLUL, glutamate transporter SLC1A2, and ornithine aminotransferase OAT (Fig. S3). Third, co-expression analysis indicates that stratifying/controlling genes in iHCC2 and their co-expressed genes are positively co-expressed with CTNNB1 (Pearson's $r > 0.32$, $Q < 0.01$, Fig. 4). This is not observed in the case of iHCC3/iHCC1 genes, which are negatively co-expressed with AKT1 or MTOR (Pearson's $r < -0.2$, $Q < 0.01$). Fourth, the association between Wnt signaling in iHCC2, and AKT activation in iHCC3 is also identified using an independent dataset of 91 HCC microarray samples and associated immunohistochemistry (Fig. 5). Associations between different HCC tumors and interferon, proliferation (PI3K/Akt activation), CTNNB1 phosphorylation/mutation (i.e. Wnt signaling), or chromosome 7 polysomy were previously identified (6). Using the authors' previously defined classes (GEO GSE9843), we observe that all tumors with CTNNB1 phosphorylating activation and mutation show high expression of iHCC2 stratifying genes. Additionally, tumors showing RPSA, AKT or IGFR activation show high expression of iHCC3

stratifying genes, thus reinforcing the relationship between PI3K/Akt/mTOR signaling activation and iHCC3. Tumor stratification based on iHCC stratifying genes shows differential distribution in the HCC subgroups previously identified (6) (Table S6). Lastly, a transcriptomic dataset with 4 HCC samples displaying CTNNB1 mutation (29) shows high expression of many of iHCC2 stratifying genes including GLUL, RHBG, SLC13A3 and ACSS3 (Fig. S6). These observations thus indicate distinct genomic features for the iHCC2 and iHCC3 phenotypes, respectively associated with aberrant Wnt signaling and PI3K/AKT/mTOR signaling activation. Interestingly, 3 stratifying genes (TDO2, KMO, XDH) and co-expressed genes (AADAT, ACMSO), are involved with the Kynurenine pathway (Fig. 4), a metabolic pathway leading to NAD⁺ production, and associated with tryptophan metabolism (30). iHCC1 also shows upregulated tryptophan metabolism in comparison with the 2 other iHCC groups (Table S4).

Together with the above observations, the observations in 3 independent datasets and considering transcriptomic, immunohistochemical, co-expression, genomic and gene-expression data (6, 9, 29, 31, 32) additionally reinforce our confidence in the newly identified stratifying genes and survival differences in iHCC1-iHCC3. Specifically, the metabolic-network derived antagonistic expression of stratifying genes identified in 186 HCC tumor transcriptomic data are consistently observed in 1. a *validation* transcriptomic dataset of 183 HCC tumors attained from TCGA (Fig. S5A); 2. a microarray dataset consisting of 221 HCC samples (Fig. S5B); 3. co-expression analysis of 369 HCC tumors from TCGA (Fig. 4); 4. a microarray dataset comprising 91 HCC tumors (Fig. 5); and 5. a comparison of CTNNB1-mutant vs noncancerous transcriptomic set (Fig. S6). Additionally, survival analysis performed on the validation TCGA dataset or Lee *et al.* dataset (Fig. S5) are consistent with the observed survival differences in iHCC1 > iHCC2 > iHCC3 (Fig. 2C).

Alternative metabolic differences between HCC subtypes

We then sought to identify metabolic differences between iHCC1, iHCC2 and iHCC3 at a pathway- and reaction-centered level using genome-scale metabolic models (GEMs). GEMs were generated for each cluster through MADE (33) and TIGER (34), using as input the differentially expressed genes, and considering biomass maximization as objective function. Fluxes in each of the models (Fig. 6A) are consistent with the hallmarks of cancer identified above (Fig. 2A) and expression data mapped into KEGG metabolic pathways (Table S3), as well as substantial metabolic differences between iHCC3 vs iHCC1 or iHCC2. Specifically, iHCC3 GEMs show lower fluxes in metabolism of amino acids, cofactors and coenzymes, pyruvate, fatty acid oxidation, carnitine shuttle, steroids, and oxidation phosphorylation compared to iHCC1/iHCC2, and lower in iHCC2 than iHCC1. When compared with iHCC1/iHCC2, iHCC3 shows higher glycolytic but lower TCA fluxes consistent with strong Warburg effect, as well as higher fluxes of fatty acid biosynthesis. Additionally, and in agreement with our previous observations (14), we observe that the low survival group iHCC3 relies on NADPH-dependent antioxidants (e.g. glutathione peroxidase/glutathione reductase) for H₂O₂ scavenging whereas the high survival

group iHCC1 relies on the NADPH-independent catalase. iHCC3 also displays higher fluxes in the pentose phosphate pathway fluxes, followed by iHCC2 and iHCC1.

Additionally, gene expression differences indicate that the 3 iHCC groups rely on alternative enzyme-coding genes for the same reactions or pathways (Fig. 3B and Fig. 6B). For instance, acetate is converted to acetyl CoA by acetyl CoA synthase, an enzyme coded by ACSS1-3. Whereas iHCC1 expresses ACSS2, which codes for the cytoplasmic form of the enzyme, iHCC2 and iHCC3 respectively express ACSS3 and ACSS1, genes that code for mitochondrial isoforms (Fig. S4). Cleavage of fructose-1,6-bisphosphate is catalyzed by aldolase, coded by ALDOA-C. While iHCC1 highly expresses the liver-specific ALDOB, iHCC3 highly expresses the non-specific ALDOA, whereas iHCC2 shows similar expression for both genes. Alternative utilization of pyruvate kinase is also observed, where iHCC1/iHCC2 show high expression of the liver-specific PKLR, whereas iHCC3 shows high expression of PKM. Glucokinase in iHCC1 is switched to hexokinase 2 in iHCC3 where ENO3 are substituted for ENO1. Additionally, PFKFB4, HSD17B6 and GLYATL2 in iHCC3 are switched to PFKFB1, HSD17B1 and GLYAT in iHCC1 and 2, respectively. Similar observations are identified in expression of genes that encode for aldehyde dehydrogenases (e.g. ALDH1B1, ALDH9A1, ALDH2, ALDH3A2 and ALDH3B1), among others (Fig. S4). A number of amino acid, sugar, cofactor, and hormone transporters are also differentially expressed between the 3 clusters, and in particular between iHCC3 and iHCC1/iHCC2 (Fig. 6B). These observations translate into distinct central metabolism, particularly between the high and low survival groups iHCC1 and iHCC3, while iHCC2 shares many of these properties with iHCC1. Several membrane transporters including amino acid, glucose and monosaccharide, choline, butyrate and citrate transporters also show substantial switching between iHCC1 and iHCC3.

Controlling genes are also differentially expressed between the 3 subtypes (Table S6), indicating different controllability metabolic behavior between them. For instance, in fatty acid elongation ELOVL6 is a controlling gene in iHCC2, but ELOVL5 is a controlling gene in iHCC1 and iHCC3. In glycerolipid metabolism, PLA2G12B is a controlling gene in iHCC1, PLA2G16 is a controlling gene in iHCC3, and both are controlling genes in iHCC2. In purine and pyrimidine metabolism, NME6 and NT5E are controlling genes in iHCC3, but not in iHCC1/iHCC2, which show other NME as controlling genes. In glycolysis, PKLR and BPGM are controlling genes in iHCC1/iHCC2, but PKM and PGAM1 are controlling genes in iHCC3. In histidine metabolism, NAA15 and SLC40A1 are controlling genes in iHCC1, SLC11A2 is a controlling gene in iHCC2, whereas NAA30 and SLC40A1 are controlling genes in iHCC3.

Above we showed the importance of controlling genes. Here, Synthetic lethality analysis performed for those stratifying and controlling genes that were found just in HCC networks not non-cancers. It highlights several potential therapeutic targets. Among controlling and stratifying genes, we find 8, 9 and 28 subtype-specific genes in iHCC1 to iHCC3, respectively, whose *in silico* knock-out leads to lethality in their respective subtype, but not in the others (Fig. 6C). Among these, we identify ALDOB, TDO2, and KMO in iHCC1, ACSS3, SQLE, and LIPT1 in

iHCC2, and ACSS1, ALDOA and G6PD in iHCC3. Knock-out of 12 controlling/stratifying genes simultaneously leads to lethality in iHCC1 and iHCC2 and includes IDH1, SORD and ARG1. These observations point towards several HCC-specific potential therapeutic targets that may be used to target low (iHCC3), intermediate (iHCC2), and high (iHCC1) survival groups.

Poor survival-associated genes show hypoxic behavior validated experimentally in human cancer cell lines

Several stratifying genes in the 3 iHCC groups are associated with redox metabolism (Fig. 3 and Fig. 5) such as G6PD, PKM and ALDOA in iHCC3, or ALDH2, and MTHFD1 in iHCC1 and iHCC2. Gene expression differences translate into altered redox metabolism (Fig. 2A) and antioxidant defenses (e.g. catalase or glutathione-based H_2O_2 scavenging, Table S5). Additionally, we have previously observed that stratification of HCC samples based on redox gene expression or acetate metabolism (7, 14) is associated with differential redox metabolism. Indeed, the expression of HIF1A is substantially higher in iHCC3 than in iHCC1 or iHCC2 (\log_2 fold-change ≈ 1 , $Q < 0.05$).

We performed a transcriptomic analysis of HepG2 cells grown under hypoxia and normoxia. We observe (Fig. 7A, $Q < 0.01$) that differentially expressed genes tend to be associated with responses to stress and to oxygen, NADH, ADP and RNA metabolism, immune system and tissue development, biological processes that are upregulated under hypoxia. In turn, DNA metabolism, replication and repair, and cell cycle related processes are upregulated under normoxia. Further, among the differentially expressed genes, the expression of stratifying genes PKM, ALDOA, MTHFD1L, ENO1, PDE9A and controlling genes in iHCC3, is significantly higher under hypoxic than normoxic conditions ($Q < 0.01$, Fig. 7B). Interestingly, stratifying and controlling genes in iHCC2 or iHCC1 are either unchanged or show downregulated expression under hypoxia (Fig. 7C and D, respectively). Additionally, among the 28 controlling genes exclusive to iHCC3 (Fig. 6C), we find that the expression of OCRL, PTPN12, HPSE, ACLY, LPCAT1, RRM2, and SPTLC2 is upregulated under hypoxia (Fig. 7B). Among those stratifying/controlling genes upregulated under hypoxia in iHCC3, we find multiple upregulated biological processes that also involve these genes, including those involved in energetic, carbohydrate, and nucleotide metabolism, and tissue development (Fig. 7E). These observations indicate that stratifying and controlling genes in iHCC3 and iHCC1/iHCC2 tumors respond strongly and antagonistically to hypoxia.

DISCUSSION

Given the high heterogeneity in cancer, many others have tried to stratify HCC patients through unsupervised clustering of tumors based on genomic and transcriptomic properties (2, 3, 9). This led to the valuable discovery of many patient group-specific differences such as cholangiocarcinoma-like features traits (CCL) (2), hepatic stem-like phenotypes (3), signaling differences (4-6), recurrence risk (10, 27), and metabolism (7, 14). However, a *functional*- and *network topology*-based characterization and stratification of HCC has never been attempted. Here,

we combined objective-dependent and independent approaches to perform a *functional metabolic network-based stratification* of hundreds of patients. Our analyses combined transcriptomic, immunohistochemistry, and clinical data across 4 datasets with hundreds of HCC tumors, a cell-line experiment, and *in silico* approaches including genome-scale metabolic modeling, gene silencing, co-expression and network analysis. We identified distinct metabolic, genomic, signaling, survival and clinical properties in 3 major HCC subtypes, iHCC1 – iHCC3. The 3 iHCCs also present 18 metabolic genes highly expressed by one group but not the others, i.e. *stratifying genes*. Our analyses showed that these genes consistently stratify HCC tumors from independent datasets. We have additionally identified 32 *controlling genes*, those that display pivotal roles in controlling network, and whose targeting would lead to lethality in one of the HCC subtypes.

Tumors in the low survival and progressive (higher grade) iHCC3 group show several signatures of low-survival and High recurrence (9, 10), and high expression of markers of poor survival (11). In turn, the high and moderate survival groups iHCC1 and iHCC2 are associated with markers of low recurrence, high survival, hepatocyte differentiation and maturity (4, 10, 27). iHCC1, iHCC2 and iHCC3 respectively display high expression of genes involved in acetate metabolism, ACSS2, ACSS3, and ACSS1. These observations are consistent with our previous analyses which showed that HCC tumors stratified based on acetate metabolism display distinct survival behavior (7). ACSS2 is highly expressed by healthy liver tissue (7) consistent with the high expression displayed by the high-survival iHCC1 group. On the other hand, ACSS1 is highly expressed in iHCC3, consistent with previous observations in a low survival group (7). Finally, we here identify for the first time that iHCC2 displays high expression of ACSS3, unlike other iHCCs.

In turn, iHCC3 tumors show high expression of ACSS1 and are associated with hypoxic environment. Experiments with HepG2 cells additionally show strong and opposing responses to hypoxia by different iHCCs. Stratifying and controlling genes in iHCC3 are upregulated under hypoxia when compared with normoxia. This is opposed by those in iHCC1 and iHCC2, which are either unchanged or show decreased expression under hypoxic conditions. These observations indicate opposing hypoxic responses under low and high survival, consistent with our previous observations (14).

iHCC1 is the tumor group with the highest survival, and shows high inflammation response compared to iHCC2. Interestingly, several stratifying genes or their co-expressed genes are involved in the Kynurenine Pathway (KP). This pathway is found upstream of NAD biosynthesis, and is the main tryptophan sink in the cell (30). Several of its genes are upregulated in adipose tissue in obesity (35, 36), consistent with the observation that 56% of patients in iHCC1 are obese or overweight. Several metabolites of the KP have been associated with inflammatory and immune responses, for instance in the induction of cytokines and macrophage-induced chemokines (37, 38). Interestingly, KP overactivation has been observed in T2D and is one of the T2D-driving mechanisms observed in pre-diabetic patients (39). The observation that T2D is one of the risk factors for HCC (39), raises the possibility that the iHCC1 phenotype is associated with T2D, unlike other iHCCs. This is reinforced by the observation that genes TDO, KMO, AADAT and ACMSD and IL6R, stratifying genes in iHCC1 or their co-expressed genes, are upregulated in

obesity or T2D (39). Finally, both T2D and obesity show promoted fatty acid oxidation, similarly to iHCC1, which displayed the highest fatty acid oxidation of the 3 iHCCs, and suggesting a potential association between those diseases and iHCC1 tumors. Overall, iHCC1 showed the highest fluxes in metabolism of amino acids, cofactors and coenzymes, pyruvate, fatty acid oxidation, carnitine shuttle, steroids, TCA and oxidative phosphorylation.

iHCC2 shows higher similarity to iHCC1 than iHCC3, but also exhibits specific features including lower fatty acid biosynthesis and high glutamine metabolism, and β -catenin-associated upregulated fatty acid oxidation. One of the main features of iHCC2 tumors is the association with β -catenin pathway alterations. CTNNB1 encodes for β -catenin, which is mutated in $\approx 20\%$ HCC (40). Mutations in this gene are associated with increased concentration of nuclear β -catenin and its target genes (e.g. glutamine synthetase GLUL, and glutamate transporter SLC1A2), and lower patient survival (41-43). Glutamine synthetase is involved in ammonia detoxification, and β -catenin-controlled induction of GLUL leads to autophagy in HCC (44, 45). β -catenin controls mitochondrial homeostasis by regulating the citric acid cycle (TCA) and fatty acid oxidation, and protects against alcohol-induced liver injury or ethanol-induced metabolic stress (Fig. 8) (46). It is consistent with the overactivation of the pathways involved in detoxification i.e. drug and xenobiotic metabolism in comparison with other subtypes. β -catenin also regulates the expression of acetaldehyde dehydrogenases (e.g. ALDH2, ALDH3A1 and ALDH3A2) (46), thus controlling TCA fluxes, as well as *stratifying gene* Acyl-CoA oxidase (AOX1) which is involved in fatty acid β oxidation (47). Our modeling analyses additionally indicate that iHCC2 shows low fatty acid biosynthesis fluxes, consistent with the negative regulation of this process by β -catenin. Sorafenib, a drug that targets expression of liver-related Wnt-targets GLUL and leads to higher sensitization in HCC tumors with high β -catenin activation (48), thus presents as a potential drug in iHCC2 but not in the other tumor groups.

Finally, iHCC3 tumors are associated with multiple features of malignant tumors, including hypoxic behavior, epithelial to mesenchymal transition (EMT), higher fluxes in fatty acid biosynthesis, and strong Warburg effect. For instance, TGF β , HIF α and NF κ B, genes involved in hypoxic response, metastasis and malignancy, are upregulated in iHCC3 and iHCC3 shares several signature activities of metastatic tissues (49). One of the main features of this tumor group is the association with PI3K/AKT/mTOR signaling activation. It also showed downstream activation of Asparagine Synthetase (ASNS), Glycolysis, and pentose phosphate pathway by PI3K/AKT/mTOR signaling (50-52), consistent with our observation in iHCC3 (Figure 8). ASNS upregulation, observed in iHCC3, is strongly correlated with metastatic potential and overexpression of ASNS promotes metastatic progression (53). Drugs targeting the PI3K/AKT/mTOR signaling or these processes, such as L-asparaginase, rapamycin, or their analogs, thus arise as potential therapeutics for the treatment of iHCC3 but not the other iHCCs. Overall, these observations highlight distinct metabolic and signaling properties in HCC tumors that stem from the high inter-tumor heterogeneity, and are associated with patient survival. *In silico* predictions identify several subgroup-specific potential therapeutic that offer full control over the metabolic network. This prompts for utilization of the above identified mechanistic differences

between tumors, together with the predicted targets, for developing personalized treatment strategies for hepatocellular carcinoma.

METHODS

Gene expression data retrieval, processing, and validation datasets

RNA-seq gene expression data and associated metadata for 369 HCC and 50 adjacent non-cancerous liver samples were retrieved from NCI's Genome Data Commons (31) as Fragments Per Kilobase of transcript per Million mapped reads (FPKM), and metabolic genes were selected based on HMR 2.0 (12). This set was split into a development set consisting of 186 HCC and 50 noncancerous samples that was used to perform all gene-expression and network-associated analyses, and a validation set consisting of 183 HCC samples that was used to verify the predictions from the development set. One additional dataset consisting of 91 HCC microarray samples was attained from GEO GSE9843 (6) was used to identify associations between gene expression and immunohistochemistry properties of the samples.

Generation of personalized and subtype GEMs

Expression data were integrated into iHCC models to construct patient-specific GEMs using tINIT (12) and RAVEN (54). The following thresholds for gene levels were considered: no expression ($\text{FPKM} < 1$), low expression ($1 \leq \text{FPKM} < 10$), medium expression ($10 \leq \text{FPKM} < 50$), and high ($\text{FPKM} \geq 50$). To assess model feasibility, i.e. biological functionality, we considered maximization of biomass production or ATP consumption as objective functions for HCC and non-cancer models, respectively. In all subsequent analyses, 57 and 56 metabolic tasks (55) were used to test model functionality, together with objective functions, respectively for HCC and non-cancer models (12). This resulted in 180 HCC and 50 non-cancer functional models that were used to construct functional gene-gene networks (fGGN).

We used MADE (33) to generate iHCC-subtype specific GEMs using as input the gene-specific fold changes and FDR determined through DESeq (56) and using TIGER (34). This flux balance approach is formulated as a single Mixed-Integer Linear Programming (MILP) problem through the CPLEX solver. Upper and lower bounds for exchange reactions in GEMs were considered based on experimental liver data (57).

Construction of personalized directed functional gene-gene networks (fGGN) and generic fGGN

After removing currency metabolites (e.g. H_2O , CO_2 , H^+ , Pi , ATP, ADP, AMP, FADH₂, NADH), we retrieved the stoichiometric matrix (S) with size $m \times n$ with m metabolites and n reactions. The ij^{th} element of the sparse matrix represents the stoichiometric coefficient of the i^{th} metabolite in j^{th} reaction in the model (Fig. S1). Positive and negative coefficients respectively represent metabolites that are produced or consumed by the reaction. Directed fGGN were generated for

personalized GEMs considering the S matrix and the vector showing the reversibility of the reaction (vector *model.rev* in Fig. S1A, where non-null elements reflect a reversible reaction). In directed fGGN, edge direction associates source to target genes if their respective reactions use the same metabolite respectively as product and reactant. For instance (Fig. S1A), metabolite c is product of reactions r_4 and r_5 (associated with genes g_4 and g_5), and reactant of reaction r_5 , thus generating the edges $g_4 \rightarrow g_6$ and $g_5 \rightarrow g_6$. Additionally, an edge is formed between both g_4 and g_5 should the network structure require both genes to be present (i.e. AND reactions). In turn, should either gene be required (i.e. OR reactions), no edge is established between genes associated with the same reaction, such as genes g_2 and g_3 in reaction r_3 (Fig. S1A). To find these edges between gene pairs, we selected all non-null row elements in the S matrix, and reaction names and respective genes were retrieved. Directed fGGNs thus permit assessing network controllability, take into account reaction direction, exclude currency metabolites, and distinguish between multi-gene protein complexes (AND reactions) from alternative enzyme isoforms (OR reactions, e.g. PKM vs PKLR).

Centrality and Controllability of personalized fGGN

To identify central genes (nodes) based on functional metabolic network analysis, we computed betweenness centrality, normalized closeness centrality, eccentricity centrality, and degree centrality for all nodes in directed fGGN. Each of these centrality parameters quantifies how central a node is in the network. Betweenness centrality quantifies the number of times a node v is found between the shortest path connecting two other nodes. Closeness and Eccentricity centrality respectively quantify the average and maximum shortest path length between node v and all other nodes in the network. Degree centrality of a node v accounts for the number of edges formed between that node and other nodes in the network, and is given in here as the sum of node input plus output (i.e. *in degree* plus *out degree*). Random networks were constructed through the Erdős–Rényi model (Fig. S2A), and topology parameters were computed through the *igraph* R package (58). All generated personalized fGGNs showed scale free (power-law) degree distribution.

We tested for the similarity between individual sample's topology parameters in HCC and non-cancer networks against all other HCC and non-cancer networks (Fig. S2B). For each individual i that had both HCC and noncancerous fGGN, we retrieved the gene-wise vector of centrality measures, and computed the Euclidean distance between this vector and those of other samples (including HCC and noncancerous). This generates 4 different distance vectors for each individual, representing the distances between the individual's tumor h or noncancerous m samples against all other tumors H and noncancerous M samples: $v_{i,h,H}$, $v_{i,h,M}$, $v_{i,m,H}$, $v_{i,m,M}$. We then compared these distance vectors to test whether the topology parameters in HCC and noncancerous networks show higher similarity to all other HCC ($R_{i,h}$) or noncancerous ($R_{i,m}$) networks

$$R_{i,h} = \frac{v_{i,h,H}}{v_{i,h,M}}$$

$$R_{i,m} = \frac{v_{i,m,H}}{v_{i,m,M}}$$

With $v_{i,h,H} = \text{distance}(h_i \text{ vs } H)$, $v_{i,h,M} = \text{distance}(h_i \text{ vs } M)$, $v_{i,m,H} = \text{distance}(m_i \text{ vs } H)$, $v_{i,m,M} = \text{distance}(m_i \text{ vs } M)$.

Thus, if $R_{i,h} \parallel R_{i,m} < 1$ indicates higher topological similarity between the sample and other HCC networks, as opposed to $R_{i,h} \parallel R_{i,m} > 1$ that show higher topological similarity between the sample and other noncancerous networks.

We identified minimum driver set nodes (MDS) in complex networks according to the Popov–Belevitch–Hautus (PBH) rank condition (59) similarly to previous approaches (19). Briefly, the number of MDS is the maximum geometric multiplicity of the eigenvalue ($g \times L$), or $N - \text{rank}(L \times I_N - A)$ for a network with size N , eigenvalue L of the adjacency matrix A , and identity matrix I_N . Driver nodes are the linearly dependent rows of reduced row echelon form of the matrix $A - L \times I_N$. Nodes are classified as Indispensable, Neutral, and Dispensable nodes by assessing the maximum geometric multiplicity of the eigenvalue $g \times L$ upon removing nodes 1 by 1. Nodes are respectively considered indispensable, neutral or dispensable if $g \times L$ increases, does not change, or decreases.

Identification of controlling genes through *in silico* gene silencing analyses

We determined network-controlling genes by performing *in silico* gene silencing, an adaptation of gene essentiality and lethality analyses but where we considered previously identified (12) metabolic tasks to assess the effect of gene silencing (through the *checkTasks* function in RAVEN). The effect of silencing each of the 2892 metabolic genes was simulated using personalized GEMs by removing each gene-associated reaction. We then repeated the *in silico* gene silencing technique regarding the flux of the objective function for the above identified controlling genes and using subtype-specific GEMs. Because we seek to identify genes whose silencing affects HCC, but not noncancerous samples, we excluded all controlling genes simultaneously identified in HCC and noncancerous models, as well as those associated with replication, transcription and translation.

Network-based stratification of fGGN

We developed a method for tumor stratification of fGGN inspired by network-based stratification of tumor mutations (24). Briefly, the adjacency matrix A of a generic fGGN is converted to degree-normalized adjacency matrix B according to the function

$$B_{ij} = A_{ij} / \sqrt{D(i,i)D(j,j)}$$

where D is the diagonal matrix of A . We apply network propagation (smoothing) to spread the influence of gene expression on the entire network through the iterative function

$$F_{t+1} = \alpha F_t B + (1 - \alpha) F_0$$

where F_0 is a matrix $m \times n$ with m genes and n patients (Fig. S1B), α is a tuning parameter that controls the distance that expression of a gene is allowed to propagate over the network. The function was iteratively run until F_{t+1} converges, i.e. matrix norm of $F_{t+1} - F_t < 10^{-5}$. We then used gene expression data to build personalized smoothed networks, and applied quantile normalization to ensure that all patients follow the same gene-expression distribution.

Following the normalization, Non-negative matrix $F_t(m \times n)$ was utilized to stratify the patients using Nonnegative Matrix Factorization (NMF) algorithm (60, 61). To ensure robust clustering, network-regularized NMF was performed 1,000 times on subsamples of 80% of the dataset, and the factorization rank was determined by performing 500 runs for each k (from 2 to 12). Because the cophenetic method indicates that both rank $k = 2$ or 3 show similar robustness (Fig. S2D, E), and substantially higher than $k = 4$, we choose the smallest rank k for which the cophenetic correlation coefficient starts decreasing (60), i.e. $k = 3$. The high robustness showed both by $k = 2$ and 3 stems from the similarity between iHCC1 and iHCC2.

Differential expression and gene set enrichment analysis, KEGG pathway analysis

Differential expression analyses were performed based on raw counts through *DESeq* (56) using default parameters.

Gene set enrichment analysis was performed in PIANO (25) either considering manually curated gene sets for cancer hallmarks or biological processes, both retrieved from MSigDB (62).

We also performed gene expression enrichment on metabolic and signaling pathways in KEGG through Pathwave (63), with 1000 permutations. Local pathways were selected if at least three reactions (or genes) were enriched in a pathway ($Q < 0.05$).

Co-expression analysis

Co-expression analysis was performed for stratifying and controlling genes in each HCC group through TCSBN (28), and the top 25 genes co-expressed with each of the input genes were retrieved (Fig. 4). Q-values were computed from the retrieved P and considered significant if $Q < 0.01$.

Validation

To validate our observations we used the following 4 independent datasets: the RNASeq testing set from NCI; a cohort with 4 HCC and 4 noncancerous RNASeq samples obtained from GEO GSE55048 (29); the Chiang microarray dataset (6); and a 221 HCC microarray dataset with associated survival data GEO GSE14520 (9, 32).

Hypoxia experiments in HepG2 cells

HepG2 (human liver hepatoma) cells were cultured and incubated in supplemented EMEM media at 21% oxygen (normoxia) or 1% oxygen (hypoxia) for 24 hours. Total RNA was extracted HepG2

cells with the RNeasy Mini Kit (Qiagen) for sequencing. Subsequently, Single-end raw sequencing data (FASTQ files) were processed to quantify TPM and count values for transcripts by Kallisto software, using human reference (GRCh38) from ensemble release 92 (64).

Statistics

Throughout, statistic methods are indicated and considered significant after multiple hypothesis testing (Benjamin-Hochberg) where $Q < 0.05$.

FIGURES

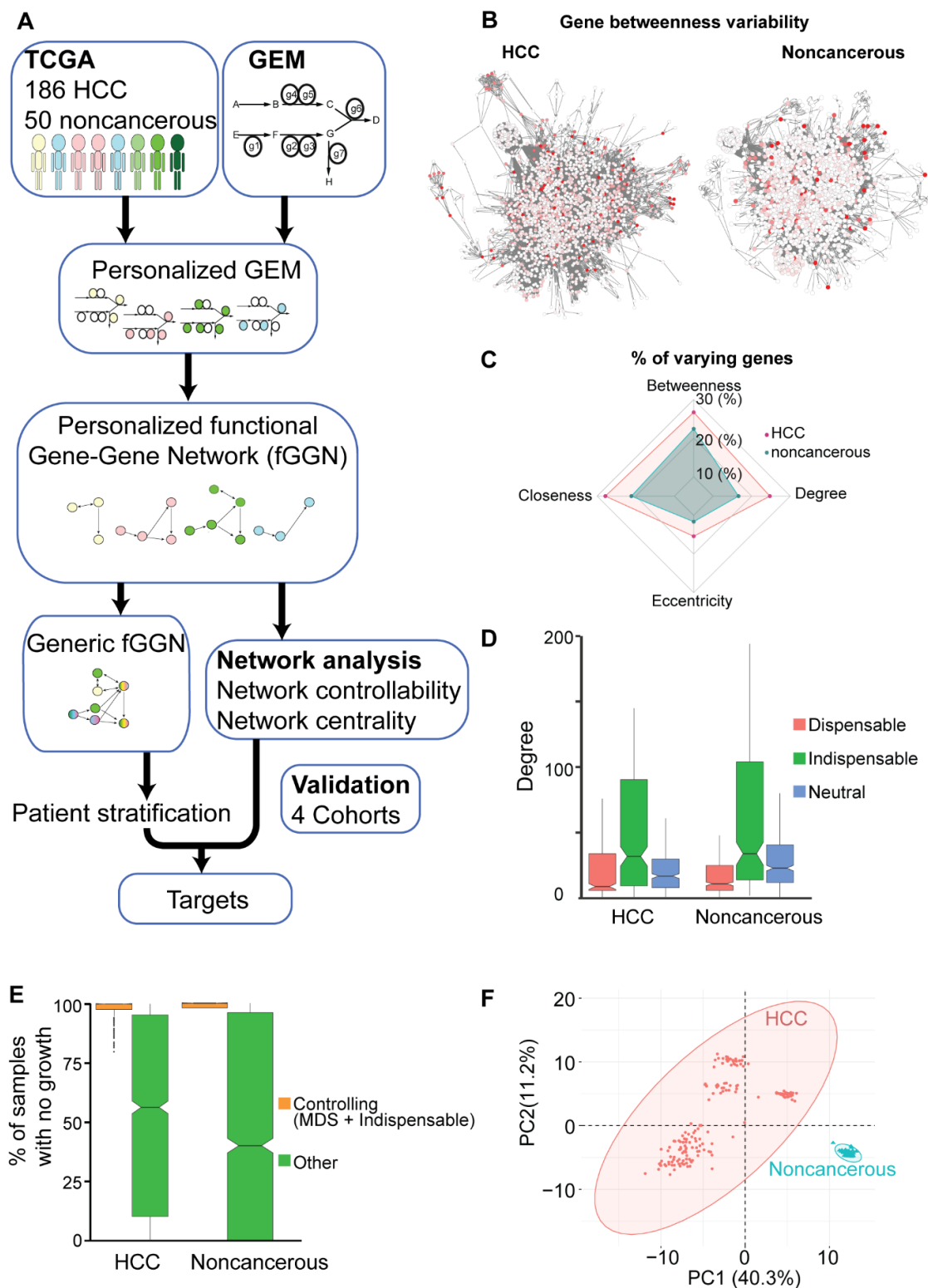
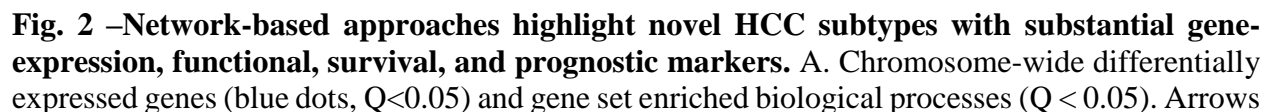


Fig. 1 – Network-based approaches to identify important genes in HCC and distinct properties from noncancerous networks. A. In our approach we integrated 186 HCC and 50 noncancerous transcriptomic samples from TCGA with genome-scale metabolic models (GEMs)

to generate personalized models integrating gene expression (circles) data. These models were then converted to *functional* gene-gene networks (fGGN) that successfully passed a series of metabolic tasks (see Methods). Then, we built a generic fGGN representative of all patients for patient stratification. We additionally used network analyses, together with patient stratification, to identify potential novel therapeutic targets for treatment. These observations were validated in 4 cohorts that included 183 HCC transcriptomic samples from TCGA (31), 91 HCC microarrays and associated immunohistochemistry (6), 8 HCC and noncancerous arrays (29), and 221 HCC tumors (9, 32). B. For HCC and noncancerous networks, 50 HCC having non-cancer expression data along with the non-cancer samples were considered. HCC and non-cancer networks show node betweenness variability. Node betweenness was computed in HCC and non-cancer networks, and the median absolute deviation was then computed within the respective network (highly varying nodes are shown as red = 1, non-varying nodes are shown as white = 0). Estimates using degree, closeness or eccentricity show similar observations (results not shown). C. Radar plot of the median node absolute deviation computed for betweenness, closeness, degree and eccentricity indicate a larger variability in the HCC vs non-cancer networks. All samples where the median absolute deviation of the four topological parameters was zero were neglected. D. the relation of topological parameter “degree” and controllability classification in both cancer and non-cancer samples. E. Silencing of controlling genes leads to lethality in >95% of the HCC (vs <50% for silencing of other genes). In noncancerous samples, silencing either kills all or none of the samples, where all controlling genes lead to lethality, versus 48% of other genes lead to lethality. E. the first two principal components of cancer and non-cancer for controllability of fGNNs, The ellipses indicate 95% confidence regions for the cancer and non-cancer samples (one outlier for non-cancer samples was observed at the 95% confidence level).



indicate direction of change (e.g. iHCC2 shows upregulated heme metabolism in comparison with iHCC1). Differential expression shows significantly differentially expressed genes for each chromosome, and horizontal lines indicate Log_2 fold changes between subtypes. B. Among all prognostic genes, we identify 42 unfavorable and 63 favorable prognostic metabolic genes differentially expressed between low and high survival groups iHCC3 and iHCC1 ($Q < 0.05$, $\text{LFC} > 1$). Prognostic markers were identified from the (11). C. Kaplan-Meyer survival analysis shows significant differences in patient survival between the 3 HCC subtypes (iHCC1 > iHCC2 > iHCC3). D. Correlation plot between tumors and mean gene expression in iHCC1 and iHCC3 ($Q < 0.01$) shows that iHCC2 tumors tend to be more similar to iHCC1 than iHCC3.

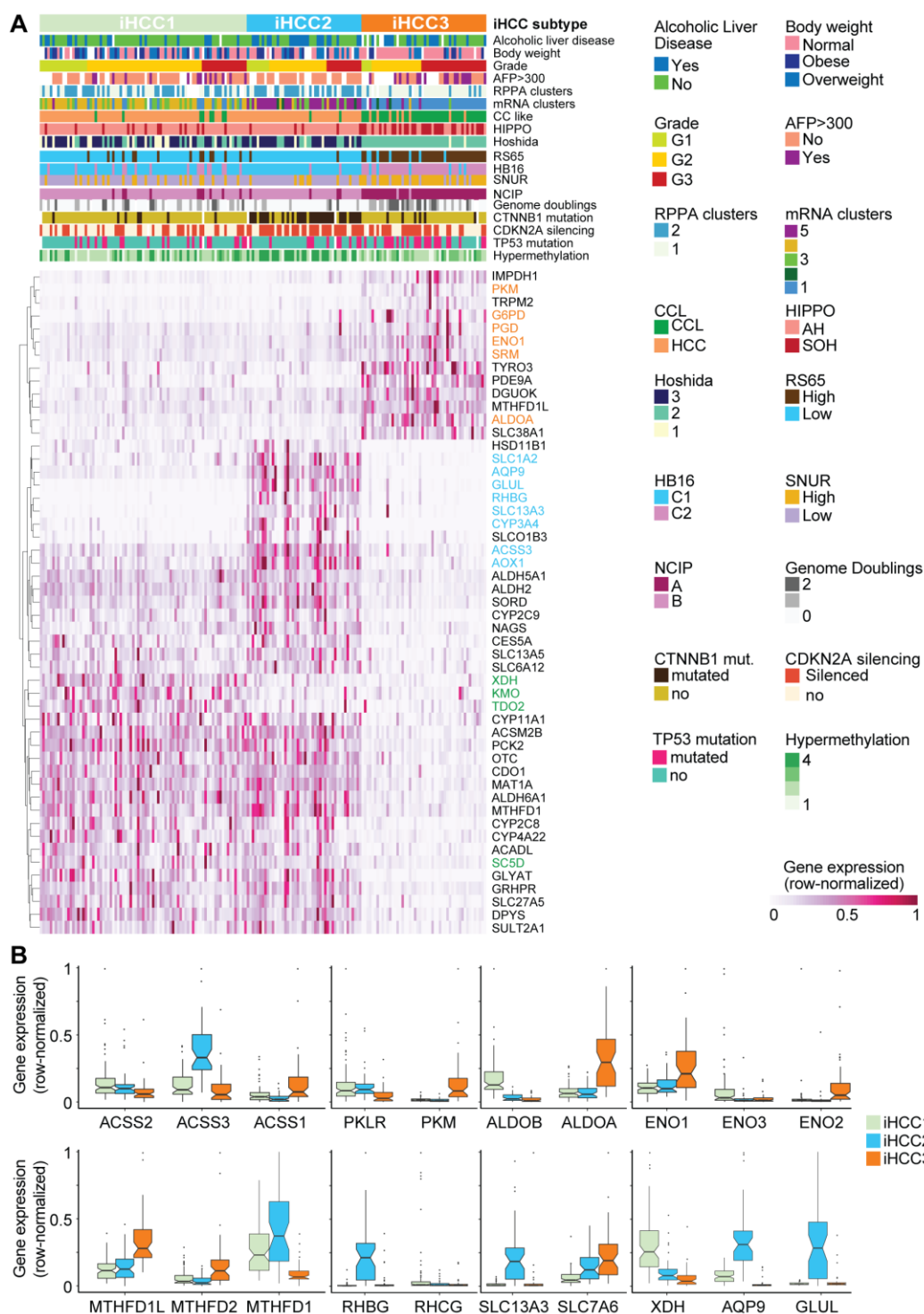


Fig. 3 – HCC tumors stratified based on metabolic network analysis show substantial gene expression, clinical, and genomic differences. A. We determined 3 novel HCC groups, and their stratifying genes are highlighted in iHCC1 (green), iHCC2 (cyan), and iHCC3 (orange). B. Expression of stratifying genes and genes that encode for enzymes catalyzing the same reactions in the 3 iHCC groups.

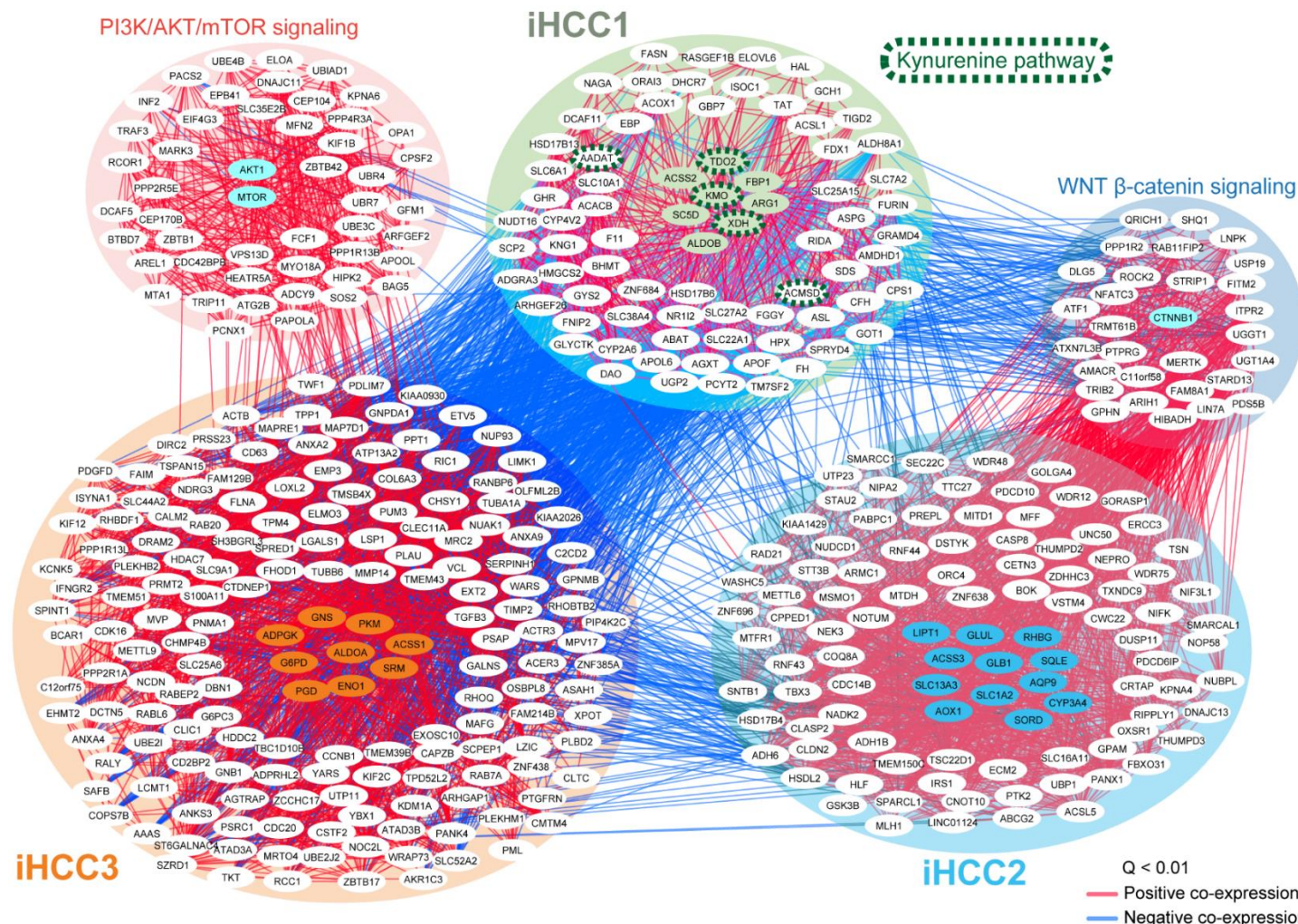


Fig. 4 – Co-expression analysis highlights association between stratifying and controlling genes in iHCC subtypes, and PI3K/AKT/mTOR or WNT/β-catenin signaling or KP pathway.s Stratifying or controlling genes for iHCC1, iHCC2 and iHCC3 were considered, and their top 25 co-expressed genes, as well as co-expression between iHCCs, were determined based on transcriptomic data in 369 HCC samples (28). We additionally considered AKT1 and MTOR, transcription factors involved in PI3K/AKT/mTOR signaling, and CTNNB1, which encodes for the transcription factor β-catenin in Wnt signaling pathway. Edges indicate positive (red) or negative (blue) Pearson correlations ($Q < 0.01$).

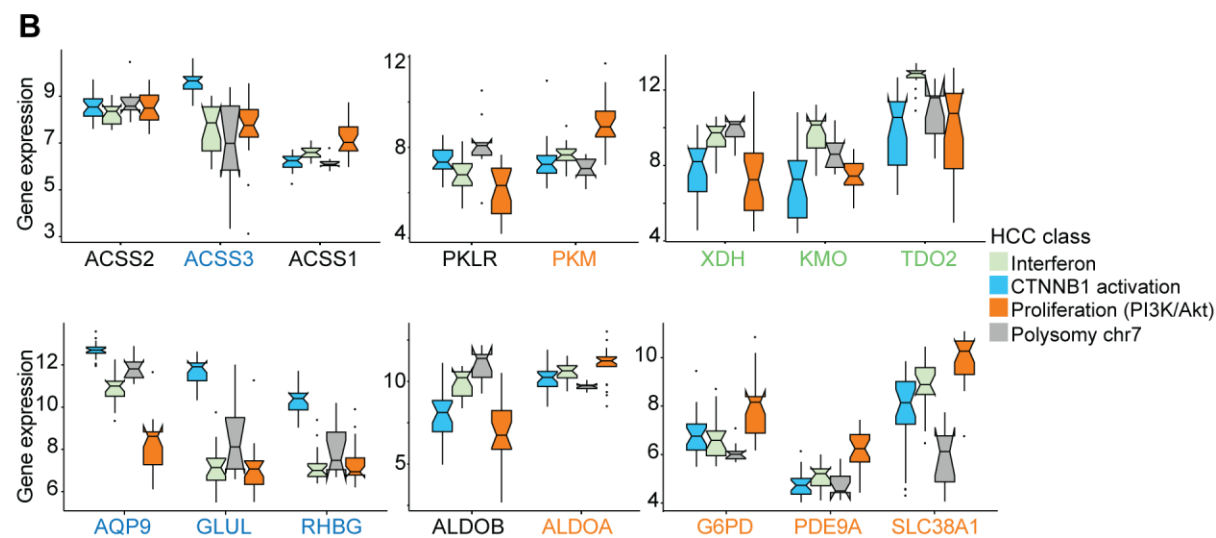


Fig. 5 – iHCC subtypes are associated with Inflammation, Wnt/ β -catenin and PI3K/AKT signaling. Gene expression data and associated immunohistochemistry were retrieved from ref.

(6) (GEO GSE9843) and stratified according to stratifying genes in iHCC (Fig. 3). A. Heatmap shows association between previously identified subclasses and iHCC subgroups identified in this work (Interferon and iHCC1; CTNNB1 activation/mutation in Wnt/ β -catenin pathway and iHCC2; PI3K/Akt signaling activation and iHCC3, Table S6). B. Stratifying genes and genes that encode for enzymes catalyzing the same reactions in the iHCC groups show similar expression patterns to those observed using transcriptomic datasets from TCGA (compare with Fig. 3B).

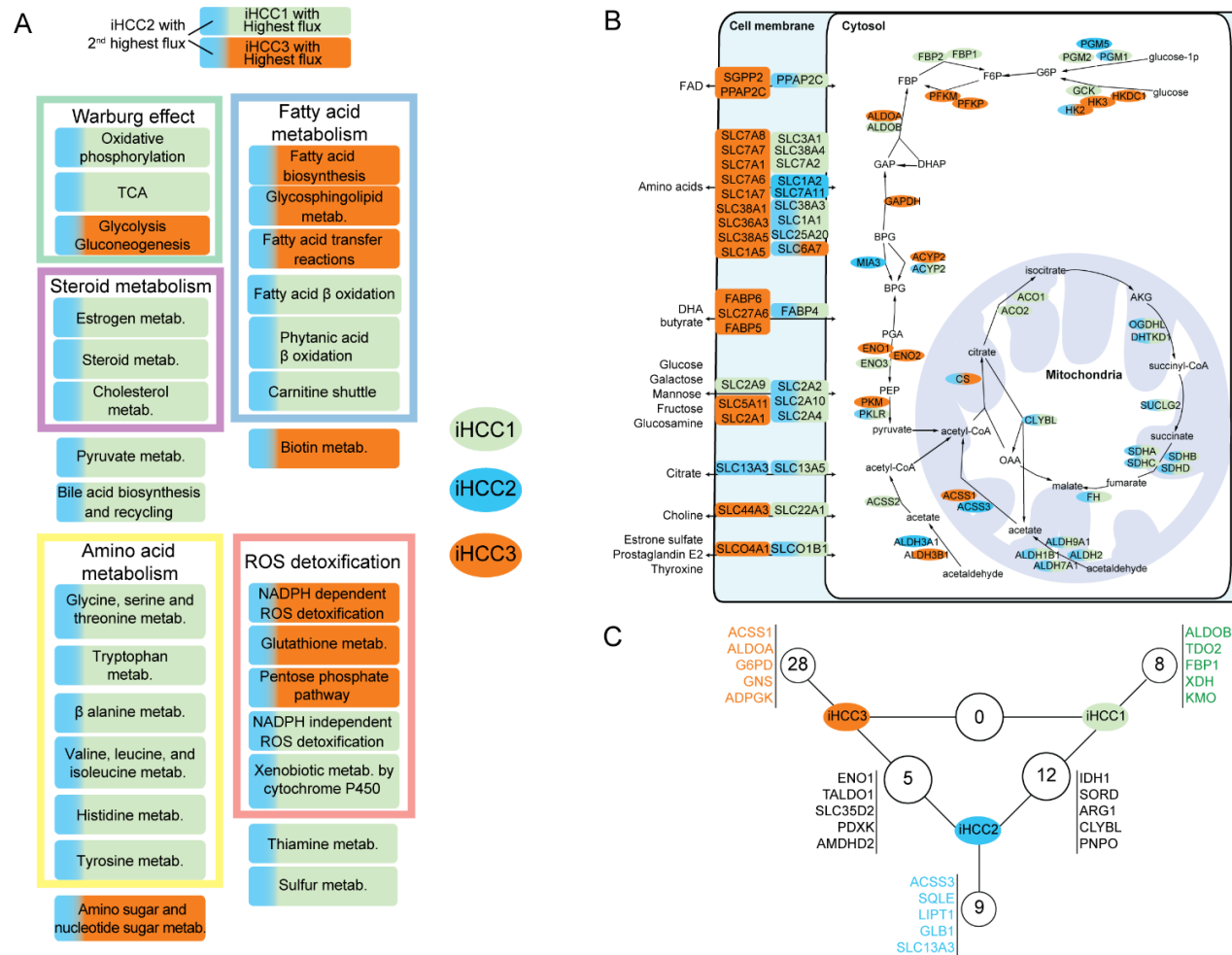


Fig. 6 – iHCC subgroups rely on alternative enzyme-coding genes for the reactions and pathways, but show specific synthetic lethal genes. A. Flux balance analysis performed on iHCC subgroup-specific models shows that iHCC1 or iHCC3 display the highest pathway reaction fluxes, followed by iHCC2. Predominant color in each box shows the iHCC subgroup that displays the highest flux. B. Metabolic genes involved in transport, Glycolysis and TCA are shown, colored according to expression in each of the subgroups. C. Number of synthetic lethal genes found in iHCC subgroups are shown. For each subgroup, 5 of those synthetic lethal genes are highlighted. Note that no synthetic lethal genes are simultaneously identified in iHCC1 and iHCC3, but several are found between iHCC2 and the two other subgroups.

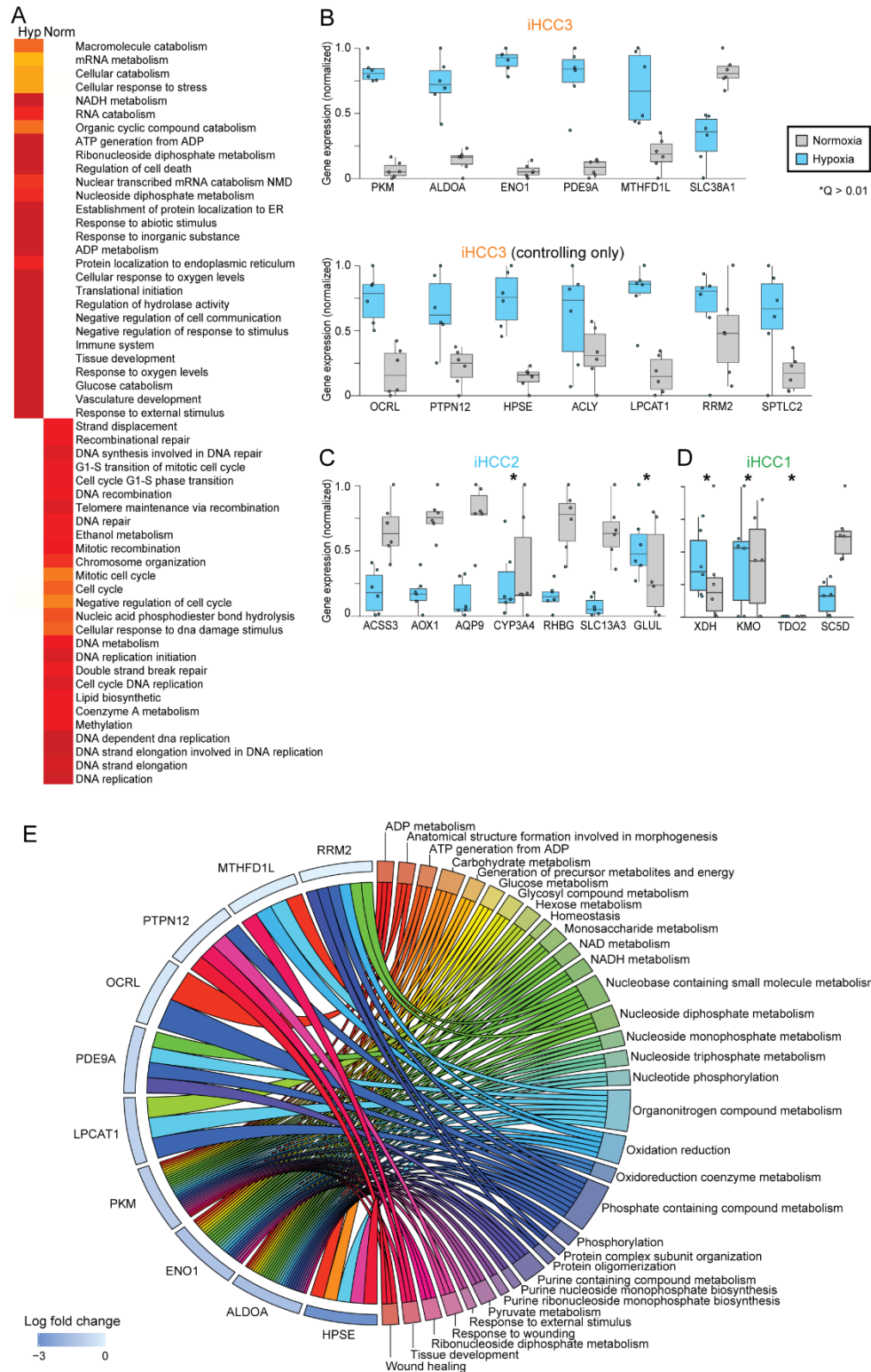


Fig. 7 –Stratifying and controlling genes in iHCC3 show specific responses to hypoxia. HepG2 cell lines were grown under normoxic or hypoxic conditions (n = 6 per condition) and mRNA was quantified. Gene set enrichment analysis was performed for differentially expressed genes in hypoxia (Hyp) vs normoxia (Norm) (A). Expression of stratifying and

controlling genes in iHCC3 (B), iHCC2 (C), and iHCC1 (D). Middle row shows only controlling genes. All genes with exception to CYP3A4, GLUL, XDH, KMO, and TDO2 are differentially expressed between hypoxia vs normoxia ($Q < 0.01$).

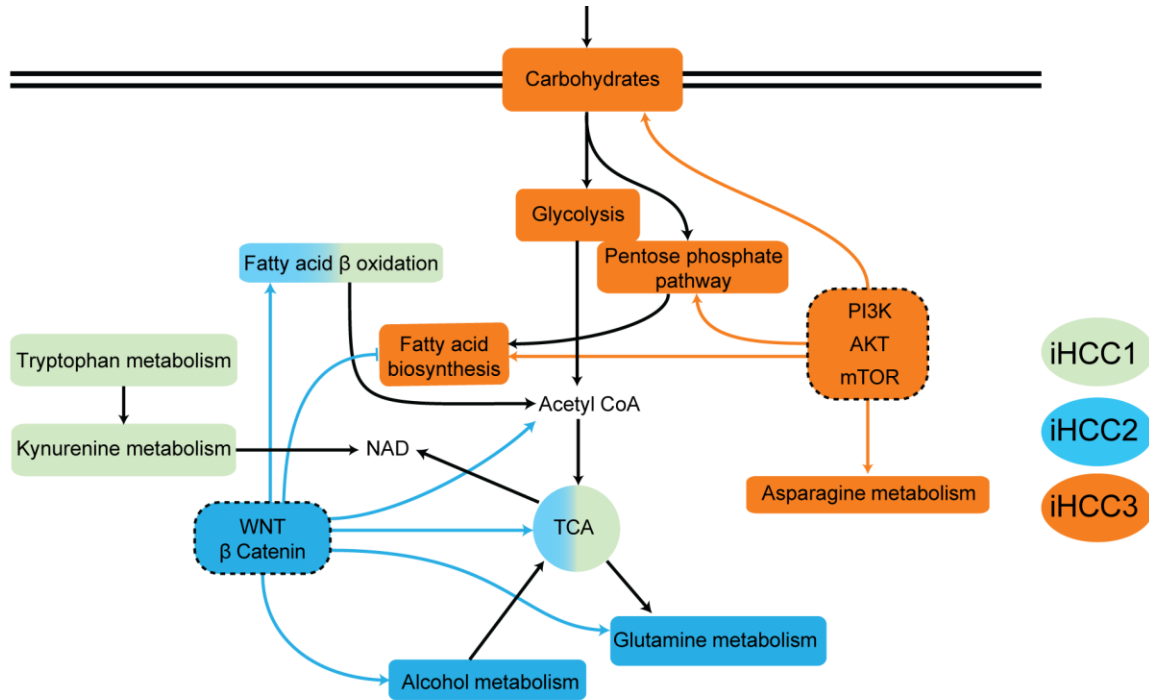


Fig. 8 – Schematic representation of the main metabolic features in iHCC1, iHCC2 and iHCC3. The main metabolic and signaling alterations are shown for iHCC1, iHCC2 and iHCC3. Boxes with 2 colors are altered in iHCC1 and iHCC2. Dotted boxes indicate altered signaling processes, and colored arrows indicate their effect on metabolism.

AUTHOR CONTRIBUTIONS

Conceptualization and design, GB and AM; Algorithm and resources, GB; Formal analysis and curation, GB and RB; Experimental validation, MK; RNA-Seq analysis, GB; Initial draft, GB, RB; Revisions & editing, All authors; Supervision, AM.

FUNDING

This work was financially supported by the Knut and Alice Wallenberg Foundation.

REFERENCES

1. Ferlay J, *et al.* (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 127(12):2893-2917.

2. Woo HG, *et al.* (2010) Identification of a cholangiocarcinoma-like gene expression trait in hepatocellular carcinoma. *Cancer Res* 70(8):3034-3041.
3. Cairo S, *et al.* (2008) Hepatic stem-like phenotype and interplay of Wnt/beta-catenin and Myc signaling in aggressive childhood liver cancer. *Cancer Cell* 14(6):471-484.
4. Hoshida Y, *et al.* (2009) Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res* 69(18):7385-7392.
5. Sohn BH, *et al.* (2016) Inactivation of Hippo Pathway Is Significantly Associated with Poor Prognosis in Hepatocellular Carcinoma. *Clin Cancer Res* 22(5):1256-1264.
6. Chiang DY, *et al.* (2008) Focal gains of VEGFA and molecular classification of hepatocellular carcinoma. *Cancer Res* 68(16):6779-6788.
7. Bjornson E, *et al.* (2015) Stratification of Hepatocellular Carcinoma Patients Based on Acetate Utilization. *Cell Rep* 13(9):2014-2026.
8. Cancer Genome Atlas Research Network (2017) Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* 169(7):1327-1341 e1323.
9. Lee JS, *et al.* (2004) Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. *Hepatology* 40(3):667-676.
10. Woo HG, *et al.* (2008) Gene expression-based recurrence prediction of hepatitis B virus-related human hepatocellular carcinoma. *Clin Cancer Res* 14(7):2056-2064.
11. Uhlen M, *et al.* (2017) A pathology atlas of the human cancer transcriptome. *Science* 357(6352):eaan2507.
12. Agren R, *et al.* (2014) Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol Syst Biol* 10(3):721.
13. Folger O, *et al.* (2011) Predicting selective drug targets in cancer through metabolic networks. *Mol Syst Biol* 7:501.
14. Benfeitas R, *et al.* (submitted) Network analysis reveals heterogeneous response of redox metabolism in hepatocellular carcinoma patients.
15. Lv W, *et al.* (2016) The drug target genes show higher evolutionary conservation than non-target genes. *Oncotarget* 7(4):4961-4971.
16. Guney E, Menche J, Vidal M, & Barabasi AL (2016) Network-based in silico drug efficacy screening. *Nat Commun* 7:10331.
17. Barabasi AL, Gulbahce N, & Loscalzo J (2011) Network medicine: a network-based approach to human disease. *Nat Rev Genet* 12(1):56-68.
18. Vinayagam A, *et al.* (2016) Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets. *Proc Natl Acad Sci U S A* 113(18):4976-4981.
19. Yuan Z, Zhao C, Di Z, Wang WX, & Lai YC (2013) Exact controllability of complex networks. *Nat Commun* 4:2447.
20. Jeong H, Mason SP, Barabasi AL, & Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* 411(6833):41-42.
21. Yu H, *et al.* (2008) High-quality binary protein interaction map of the yeast interactome network. *Science* 322(5898):104-110.
22. Najafi A, Bidkhorl G, Bozorgmehr JH, Koch I, & Masoudi-Nejad A (2014) Genome scale modeling in systems biology: algorithms and resources. *Curr Genomics* 15(2):130-159.
23. Patil KR & Nielsen J (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci U S A* 102(8):2685-2689.
24. Hofree M, Shen JP, Carter H, Gross A, & Ideker T (2013) Network-based stratification of tumor mutations. *Nat Methods* 10(11):1108-1115.
25. Varem0 L, Nielsen J, & Nookaew I (2013) Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res* 41(8):4378-4391.
26. Yoshihara K, *et al.* (2013) Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 4:2612.
27. Kim SM, *et al.* (2012) Sixty-five gene-based risk score classifier predicts overall survival in hepatocellular carcinoma. *Hepatology* 55(5):1443-1452.
28. Lee S, *et al.* (2018) TCSBN: a database of tissue and cancer specific biological networks. *Nucleic Acids Res* 46(D1):D595-D600.

29. Ding X, *et al.* (2014) Transcriptomic characterization of hepatocellular carcinoma with CTNNB1 mutation. *Plos One* 9(5):e95307.
30. Badawy AA (2017) Kynurenine Pathway of Tryptophan Metabolism: Regulatory and Functional Aspects. *Int J Tryptophan Res* 10:1178646917691938.
31. Weinstein JN, *et al.* (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 45(10):1113-1120.
32. Lee JS, *et al.* (2006) A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells. *Nat Med* 12(4):410-416.
33. Jensen PA & Papin JA (2011) Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics* 27(4):541-547.
34. Jensen PA, Lutz KA, & Papin JA (2011) TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. *Bmc Syst Biol* 5:147.
35. Favennec M, *et al.* (2015) The kynurenine pathway is activated in human obesity and shifted toward kynurenine monooxygenase activation. *Obesity* 23(10):2066-2074.
36. Moriya C & Satoh H (2016) Teneligliptin Decreases Uric Acid Levels by Reducing Xanthine Dehydrogenase Expression in White Adipose Tissue of Male Wistar Rats. *J Diabetes Res* 2016:3201534.
37. Heng B, *et al.* (2016) Understanding the role of the kynurenine pathway in human breast cancer immunobiology. *Oncotarget* 7(6):6506-6520.
38. Opitz CA, *et al.* (2011) An endogenous tumour-promoting ligand of the human aryl hydrocarbon receptor. *Nature* 478(7368):197-203.
39. Oxenkrug GF (2015) Increased Plasma Levels of Xanthurenic and Kynurenic Acids in Type 2 Diabetes. *Mol Neurobiol* 52(2):805-810.
40. Forbes SA, *et al.* (2017) COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* 45(D1):D777-D783.
41. Okabe H, *et al.* (2016) Diverse Basis of beta-Catenin Activation in Human Hepatocellular Carcinoma: Implications in Biology and Prognosis. *Plos One* 11(4):e0152695.
42. Kim YD, *et al.* (2008) Genetic alterations of Wnt signaling pathway-associated genes in hepatocellular carcinoma. *J Gastroenterol Hepatol* 23(1):110-118.
43. Zucman-Rossi J, *et al.* (2007) Differential effects of inactivated Axin1 and activated beta-catenin mutations in human hepatocellular carcinomas. *Oncogene* 26(5):774-780.
44. Sohn BH, Park IY, Shin JH, Yim SY, & Lee JS (2018) Glutamine synthetase mediates sorafenib sensitivity in beta-catenin-active hepatocellular carcinoma cells. *Exp Mol Med* 50(1):e421.
45. Takigawa Y & Brown AM (2008) Wnt signaling in liver cancer. *Curr Drug Targets* 9(11):1013-1024.
46. Liu S, *et al.* (2012) beta-catenin is essential for ethanol metabolism and protection against alcohol-mediated liver steatosis in mice. *Hepatology* 55(3):931-940.
47. Lehwald N, *et al.* (2012) beta-Catenin regulates hepatic mitochondrial function and energy balance in mice. *Gastroenterology* 143(3):754-764.
48. Lachenmayer A, *et al.* (2012) Wnt-pathway activation in two molecular classes of hepatocellular carcinoma and experimental modulation by sorafenib. *Clin Cancer Res* 18(18):4997-5007.
49. Robinson DR, *et al.* (2017) Integrative clinical genomics of metastatic cancer. *Nature* 548(7667):297-303.
50. Duvel K, *et al.* (2010) Activation of a metabolic gene regulatory network downstream of mTOR complex 1. *Mol Cell* 39(2):171-183.
51. Jiang P, Du W, & Wu M (2014) Regulation of the pentose phosphate pathway in cancer. *Protein Cell* 5(8):592-602.
52. Toda K, *et al.* (2016) Metabolic Alterations Caused by KRAS Mutations in Colorectal Cancer Contribute to Cell Adaptation to Glutamine Depletion by Upregulation of Asparagine Synthetase. *Neoplasia* 18(11):654-665.
53. Knott SRV, *et al.* (2018) Asparagine bioavailability governs metastasis in a model of breast cancer. *Nature* 554(7692):378-381.
54. Agren R, *et al.* (2013) The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Comput Biol* 9(3):e1002980.
55. Agren R, *et al.* (2012) Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput Biol* 8(5):e1002518.

56. Anders S & Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106.
57. Hyotylainen T, *et al.* (2016) Genome-scale study reveals reduced metabolic adaptability in patients with non-alcoholic fatty liver disease. *Nat Commun* 7:8994.
58. Csardi G & Nepusz T (2006) The igraph software package for complex network research. *InterJournal, Complex Systems* 1695(5):1-9.
59. Sontag ED (2013) *Mathematical control theory: deterministic finite dimensional systems* (Springer Science & Business Media).
60. Brunet JP, Tamayo P, Golub TR, & Mesirov JP (2004) Metagenes and molecular pattern discovery using matrix factorization. *P Natl Acad Sci USA* 101(12):4164-4169.
61. Gaujoux R & Seoighe C (2010) A flexible R package for nonnegative matrix factorization. *Bmc Bioinformatics* 11.
62. Liberzon A, *et al.* (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 1(6):417-425.
63. Schramm G, *et al.* (2010) PathWave: discovering patterns of differentially regulated enzymes in metabolic pathways. *Bioinformatics* 26(9):1225-1231.
64. Zerbino DR, *et al.* (2018) Ensembl 2018. *Nucleic Acids Res* 46(D1):D754-D761.



Supplementary Information for

Stratification of hepatocellular carcinoma tumors based on metabolic network heterogeneity

Gholamreza Bidkhor, Rui Benfeitas, Martina Klevstig, Jens Nielsen, Mathias Uhlen, Jan Boren, Adil Mardinoglu

Adil Mardinoglu (adilm@scilifelab.se)

This PDF file includes:

Captions for Tables S1 – S7

Figs. S1 to S10

Other supplementary materials for this manuscript include the following:

Tables S1 to S7

SUPPLEMENTARY TABLE CAPTIONS

Table S1 – Centrality measures in HCC and noncancerous fGGNs.

Table S2 – Gene MDS and dispensability classification, and *in silico* gene silencing in HCC and noncancerous fGGNs. Number of MDS and indispensable genes is also shown for each subsystem.

Table S3 – HCC clusters of fGGN identified through Nonnegative Matrix Factorization. Differentially expressed genes between iHCC1, iHCC2, iHCC3 and noncancerous samples are indicated. Table also shows comparison of clinical properties, immune signatures, hallmarks of cancer, genes associated with prognostic genes and liver-specific genes, between iHCC1, iHCC2 and iHCC3.

Table S4 – PIANO output between clusters.

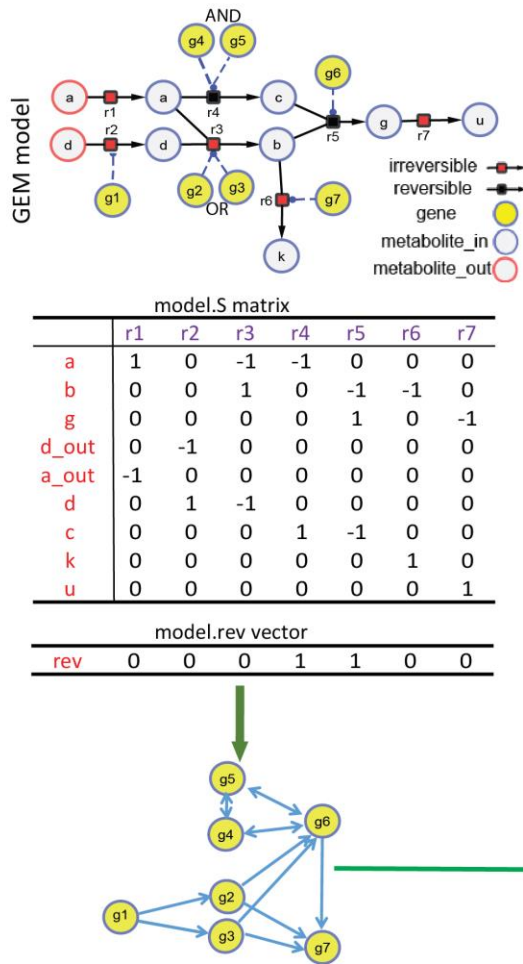
Table S5 – FBA result for comparison between iHCC1, iHCC2 and iHCC3.

Table S6 – Comparison of HCC subtypes identified by ref. (6).

Table S7 – Subtype-specific controlling genes.

SUPPLEMENTARY FIGURES

A



B

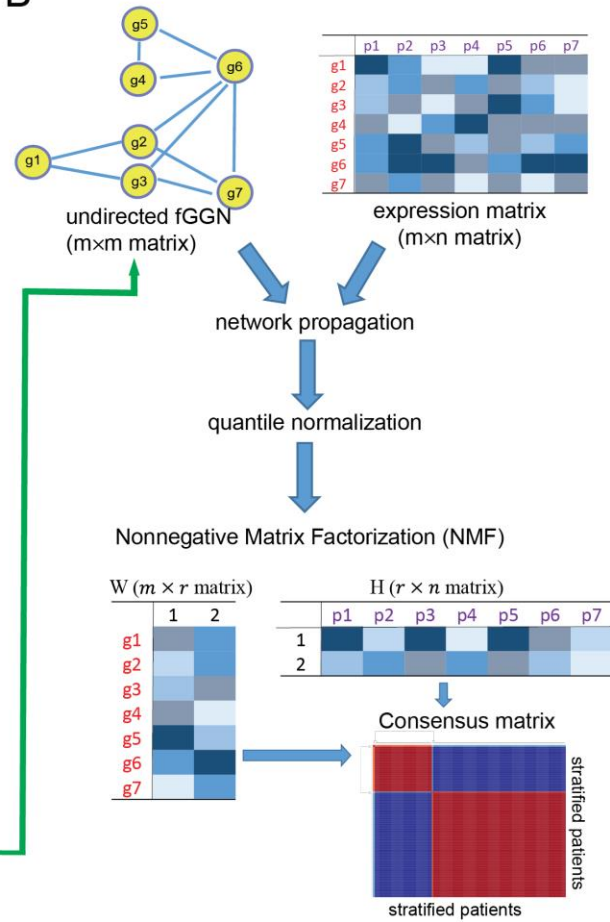


Fig. S1 – Development of a new algorithm for studying metabolic heterogeneity in HCC, and for patient stratification. A. In this method, we used functional gene-gene (fGGN) networks, generated from the stoichiometric matrix (S) and reversibility vector (rev). In fGGNs two genes (enzymes) are connected not only if they are involved in the same reaction but also if their reactions share a metabolite. B. Patient expression data was integrated with undirected fGGN, and using network propagation to generate fGGN representative of all patients' gene expression. After quantile normalization, we used Nonnegative Matrix Factorization (NMF) to stratify the patients.

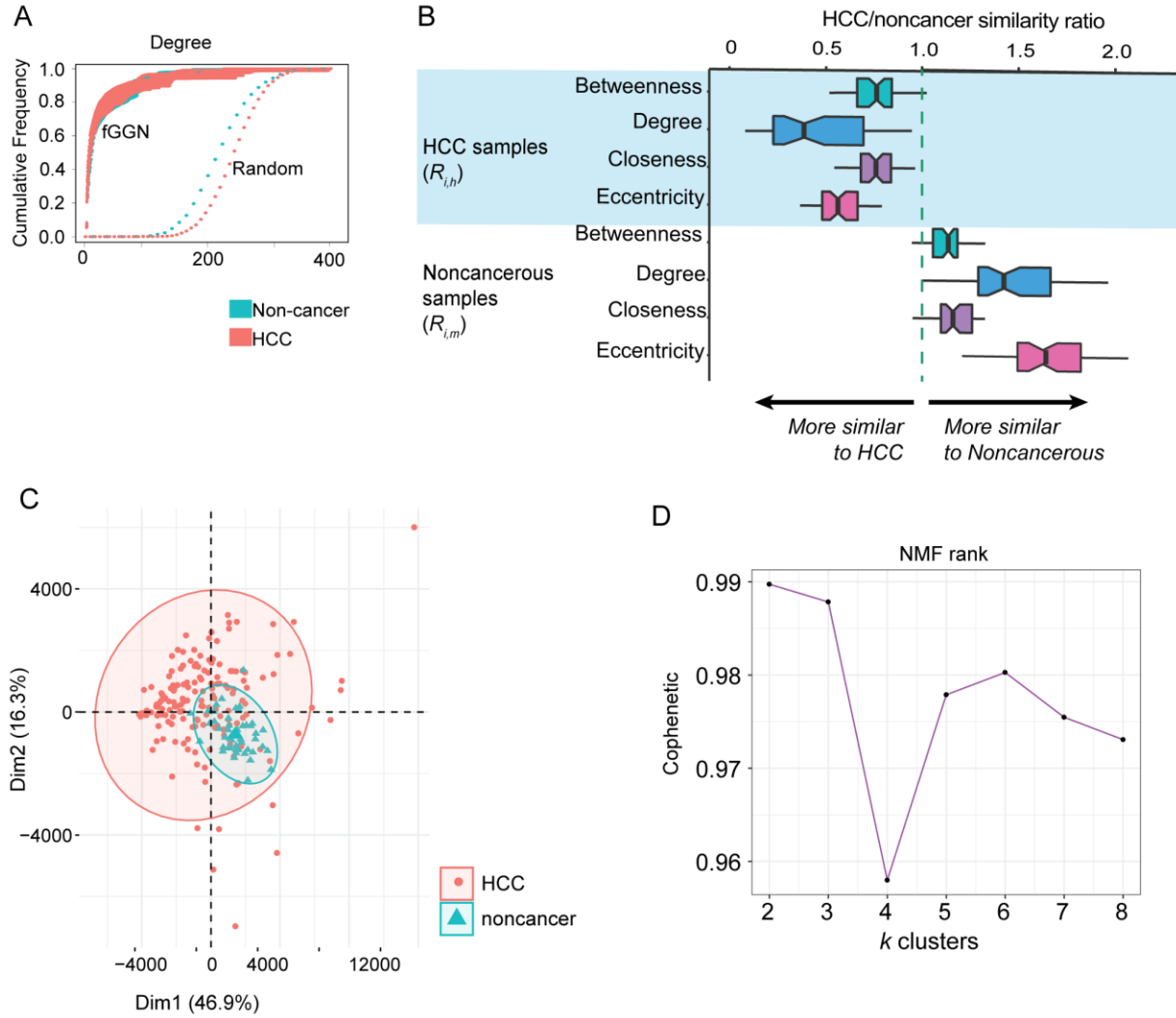


Fig. S2 – General properties in fGGN networks. A. Cumulative frequency distribution of degree in HCC and non-cancer networks, and HCC and noncancerous networks randomly generated through the Erdős–Rényi model. B. Comparison of topology parameters for HCC and noncancerous samples. For each personalized fGGN, we computed the ratio of distances of centrality measures against all other HCC and noncancer fGGN (see Methods). $R_{i,h}$ and $R_{i,m}$ respectively represent the similarity ratios HCC/noncancer (see Methods). Here, a ratio >1 shows topology features resembling noncancerous samples, whereas ratio <1 show higher similarity to HCC samples. C. Principal component analysis of gene expression data in HCC and noncancerous samples shows large overlap between the two sample groups. D. Cophenetic correlation coefficient between $k = 2 - 8$ indicates that 2 or 3 clusters provide similarly robust separation between samples. This is due to the high similarity between iHCC2 and iHCC1 (see Main Text). We choose $k = 3$, the smallest rank for which the cophenetic correlation coefficient starts decreasing (60).

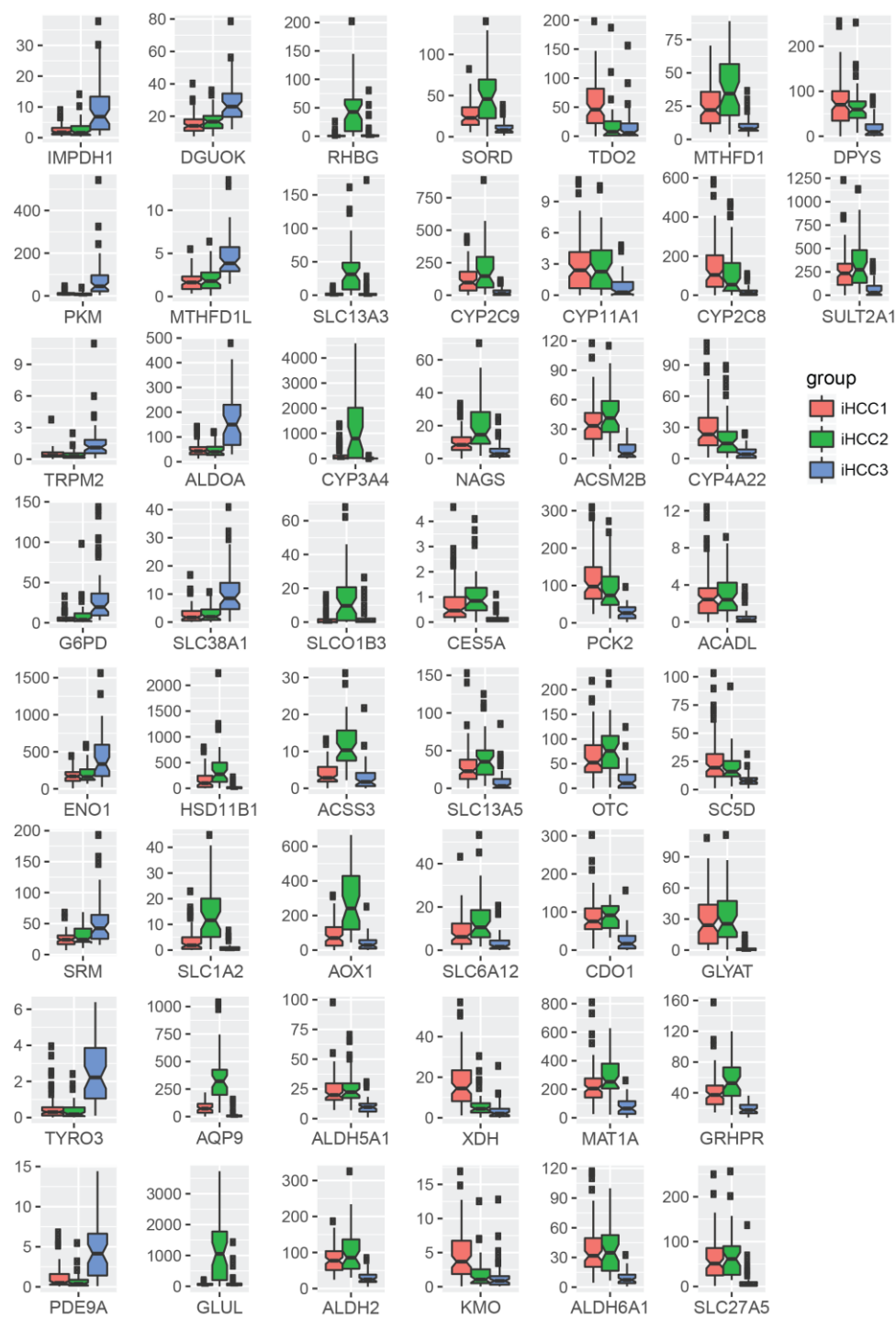
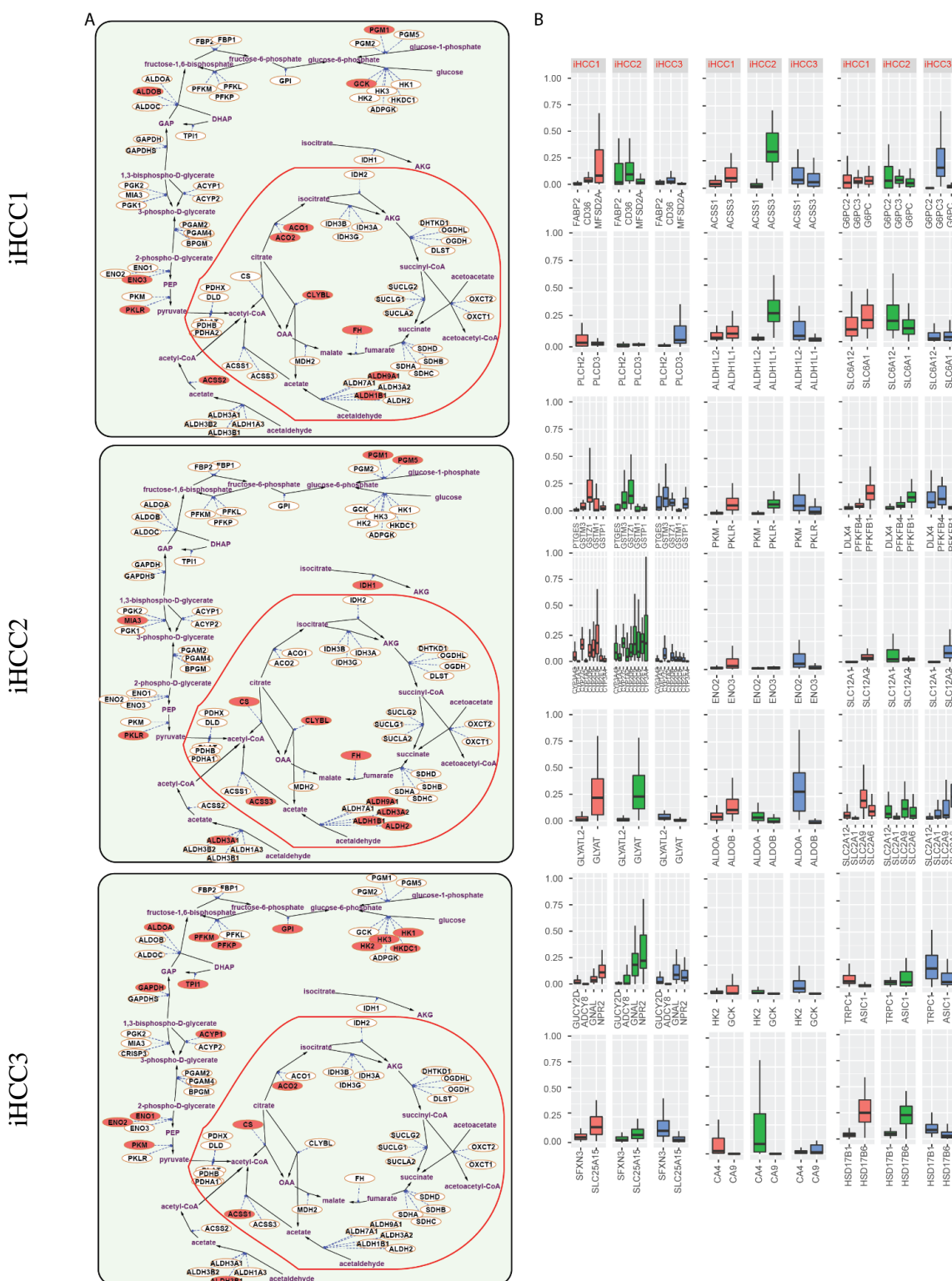


Fig. S3 – Stratifying genes in iHCC1 – iHCC3. Gene expression (FPKM) is presented for the 3 groups.



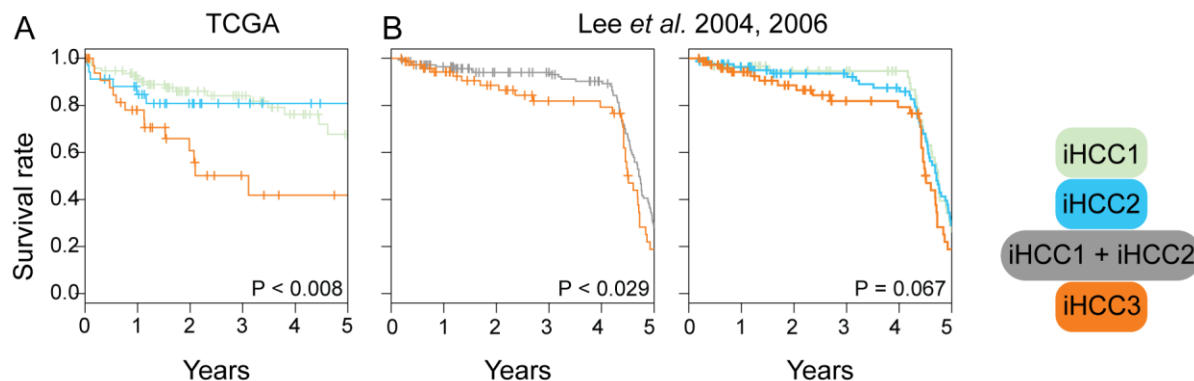


Fig. S5 – Survival differences identified in 2 independent cohorts. Using the transcriptomic validation dataset from TCGA or microarray data (32) we stratified tumors based on stratifying genes identified using network analysis. Survival analysis shows significant survival differences identified in patients associated with high expression of stratifying genes in iHCC1 – iHCC3.



Fig. S6 – A cohort with 4 HCC samples where CTNNB1 is mutated shows high expression of iHCC2 stratifying genes GLUL, RHBG, SLC13A3 and ACSS3 when compared with noncancerous samples. Transcriptomic dataset was retrieved from GEO with series id GSE55048 (29).

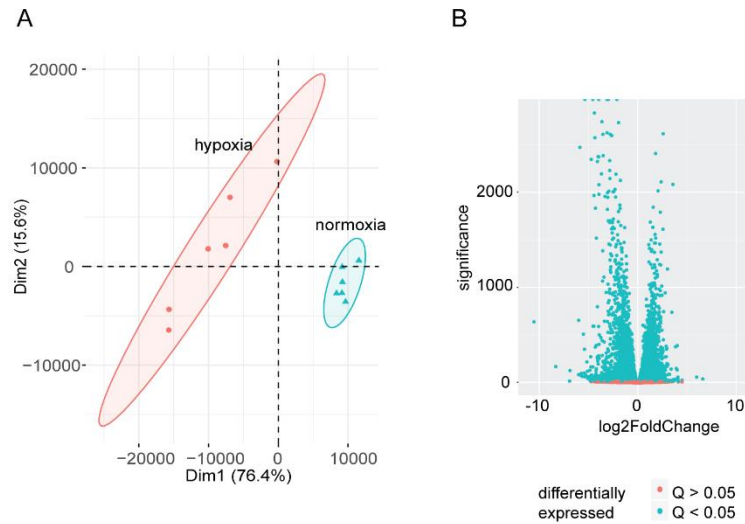


Fig. S7 – Transcriptomic data in HepG2 cell lines grown under hypoxia or normoxia are clearly distinct. HepG2 cells were grown under 21% oxygen (normoxia) or 1% oxygen (hypoxia) and RNA was extracted. PCA of gene expression (A) and differential expression analysis (B) shows substantial differences between both groups.

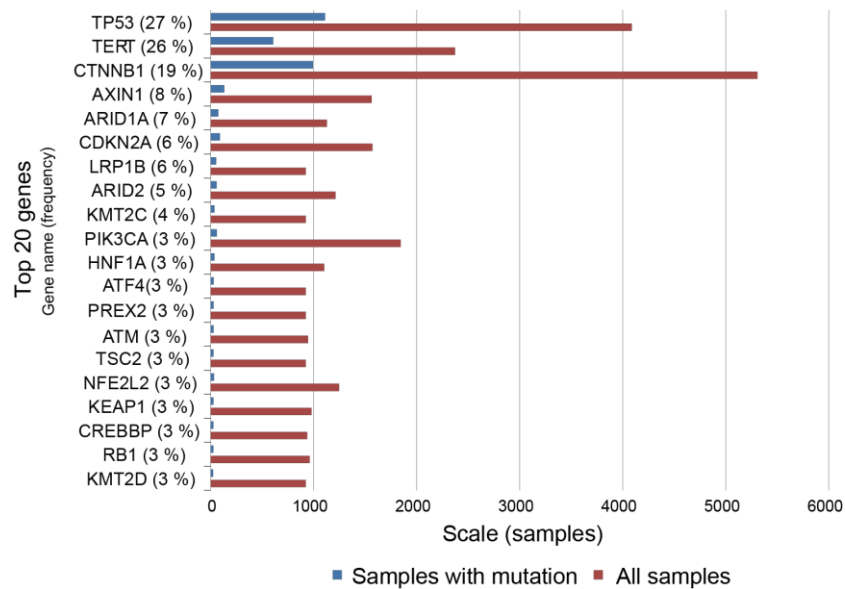


Fig. S8 – Top 20 mutated genes in HCC. Data was retrieved from COSMIC (40).

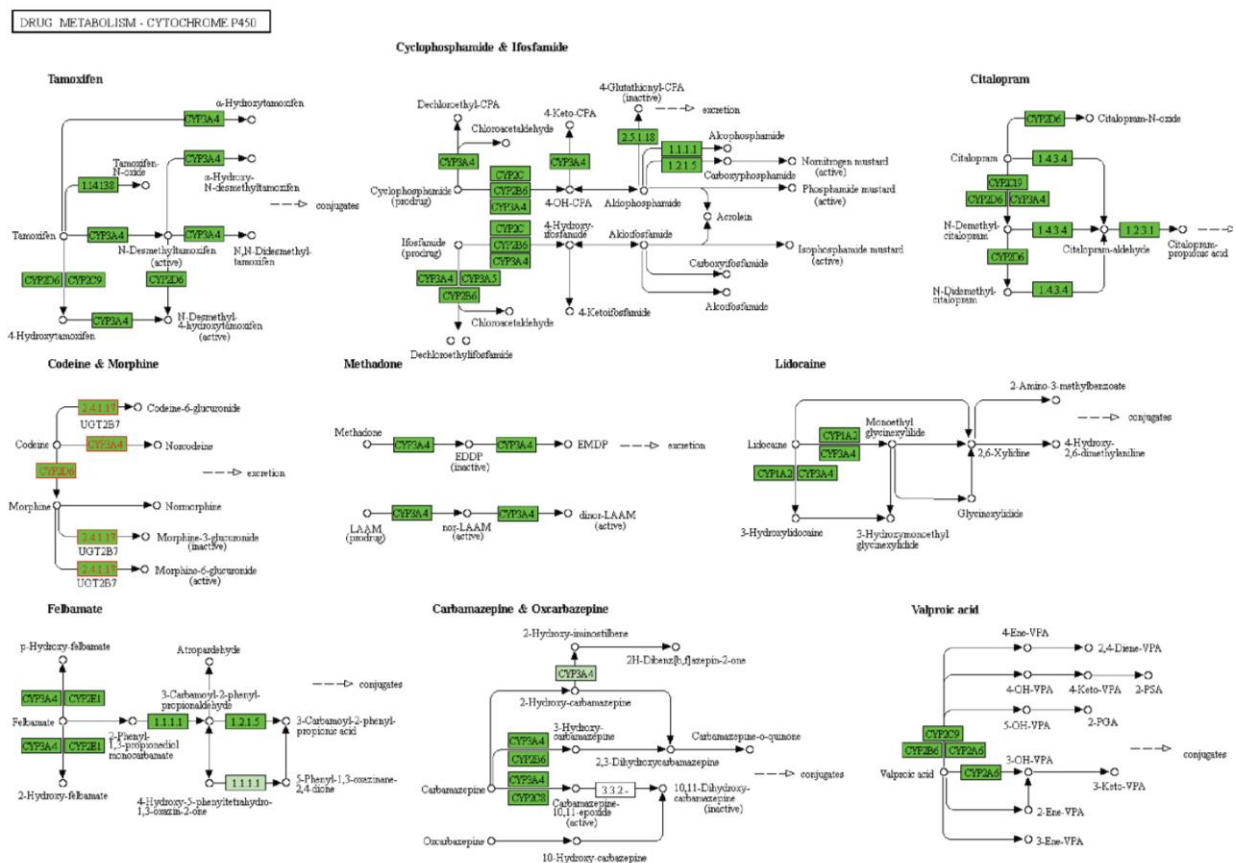


Fig. S9 – Drug metabolism shows upregulated genes (green) in iHCC2 (compared with iHCC3) as determined in Pathwave (63). No genes are downregulated in these pathways. Details in Table S6.

