ROYAL SOCIETY OPEN SCIENCE

Research

Author for correspondence:
Balázs Győrffy
e-mail: gyorffy.balazs@ttk.mta.hu

THE ROYAL SOCIETY
PUBLISHING

Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinoma

Otília Menyhárt[1,2], Ádám Nagy[1,2] and Balázs Győrffy[1,2]

[1]2nd Department of Pediatrics, Semmelweis University, H-1094 Budapest, Hungary
[2]MTA TTK Lendület Cancer Biomarker Research Group, Institute of Enzymology, Hungarian Academy of Sciences, Magyar tudósok körútja 2, H-1117 Budapest, Hungary

OM, 0000-0003-4129-4589; ÁN, 0000-0002-7979-7184;
BG, 0000-0002-5772-3766

Background: Potential prognostic biomarker candidates for hepatocellular carcinoma (HCC) are abundant, but their generalizability is unexplored. We cross-validated markers of overall survival (OS) and vascular invasion in independent datasets. *Methods:* The literature search yielded 318 genes related to survival and 52 related to vascular invasion. Validation was performed in three datasets (RNA-seq, $n = 371$; Affymetrix arrays, $n = 91$; Illumina gene chips, $n = 135$) by uni- and multivariate Cox regression and Mann–Whitney U-test, separately for Asian and Caucasian patients. *Results:* One hundred and eighty biomarkers remained significant in Asian and 128 in Caucasian subjects at $p < 0.05$. After multiple testing correction *BIRC5* ($p = 1.9 \times 10^{-10}$), *CDC20* ($p = 2.5 \times 10^{-9}$) and *PLK1* ($p = 3 \times 10^{-9}$) endured as best performing genes in Asian patients; however, none remained significant in the Caucasian cohort. In a multivariate analysis, significance was reached by stage ($p = 0.0018$) and expression of *CENPH* ($p = 0.0038$) and *CDK4* ($p = 0.038$). *KIF18A* was the only gene predicting vascular invasion in the Affymetrix and Illumina cohorts ($p = 0.003$ and $p = 0.025$, respectively). *Conclusion:* Overall, about half of biomarker candidates failed to retain prognostic value and none were better than stage predicting OS. Impact: Our results help to eliminate biomarkers with limited capability to predict OS and/or vascular invasion.

1. Background

In spite of tremendous efforts toward the discovery of novel prognostic or predictive biomarkers in solid tumours, less than 1% of these are

Table 1. List of Abbreviations.

AFP	alpha-fetoprotein
CCKR	gastrin and cholecystokinin receptors mediated signalling network
CIN	chromosome instability
DAVID	Database for Annotation, Visualization and Integrated Discovery
EGA	European Genome-phenome Archive
FDR	false discovery rate
FGF	fibroblast growth factor
GEO	NCBI gene expression omnibus
GnRHR	gonadotropin-releasing hormone receptor
HBV	Hepatitis B
HCC	hepatocellular carcinoma
HCV	Hepatitis C
HR	hazard rate
IAP	inhibitor of apoptosis protein
IGF	insulin-like growth factor
IHC	immunohistochemistry
LC	liver cancer
miRNA	micro RNA
OS	overall survival
PANTHER	Protein ANalysis THrough Evolutionary Relationships
PCR	polymerase chain reaction
PDGF	platelet-derived growth factor
PVTT	portal vein tumour thrombosis
RB	retinoblastoma
TCGA	The Cancer Genome Atlas
TKI	receptor tyrosine kinase inhibitor
VEGF	vascular endothelial growth factor

estimated to enter clinical practice [1]. One critical component behind the high failure rate is poor reporting of key study elements hampering the interpretability and clinical applicability of prognostic studies [2]. Another critical requirement would be the validation of the findings by independent investigators in independent datasets, as promising biomarkers should provide reproducible results when tested in external samples.

Worldwide, liver cancer (LC) is the second leading cause of cancer-related mortality, with the vast majority of cases (83%) occurring in the less developed parts of the world causing a major health crisis in Eastern and Southeastern Asia [3]. Nonetheless, LC is also on the rise in the European Union (47 000 deaths per year) with the highest incidence in Southern Europe [4], and despite decreasing death rates for all cancers combined, LC burden is growing rapidly in the USA as well [5].

Hepatocellular carcinoma (HCC) accounts for up to 90% of primary liver malignancies with a highly unfavourable prognosis due to fast growth, early hepatic metastasis and rapid multidrug resistance. Chronic liver diseases linked to hepatitis infections, diabetes mellitus and fatty liver disease are frequently in the background of HCC [6]. Disease aetiology differs across races: in Asia, endemic for HBV and HCV, most HCCs develop as a consequence of chronic viral infection. In developed countries, HCC is linked primarily to chronic liver disease caused by cirrhosis from excessive alcohol consumption, diabetes mellitus or non-alcoholic fatty liver disease [6]. As a result of multiple causal factors, HCC is one of the most heterogeneous cancers with a highly variable clinical course. Abbreviations are listed in table 1.

Most HCC patients are diagnosed at an intermediate or advanced stage, rendering only 30% suitable for a potentially curative therapy [7]. Since 2007, patients with an advanced stage HCC (with vascular invasion and extrahepatic spread) can be treated with sorafenib, a multitarget receptor tyrosine kinase

inhibitor (TKI). Sorafenib treatment is associated with increased median overall survival (OS) (7.9 versus 10.7 months) [8]. In the Asia Pacific study, 6.5 months OS was reported after sorafenib compared to 4.2 months after placebo [9]. Notably, the modest survival advantage was coupled with frequent side effects and with no increase in quality of life. Lenvatinib, an inhibitor of VEGF, FGF and PDGF receptors and also an RET and KIT inhibitor became recently approved as a first-line treatment for unresectable HCC [10]. Regorafenib (fluoro-sorafenib), another oral TKI, became approved as a second-line treatment for patients progressing on sorafenib. Regorafenib significantly improved both overall (7.8 versus 10.6 months) and progression free survival (1.5 versus 3.1 months) compared to placebo [11]. However, potent adjuvant therapies after surgery are still lacking, and only palliative care is available for patients with multiple metastases, with an estimated OS of less than three months.

HCC is characterized by extreme phenotypic and molecular heterogeneity, and molecular stratification has not yet been established [12]. An HCC can harbour 40–80 mutations, among those five to eight driver mutations per tumour [12]. Cytotoxic chemotherapies fail as ineffective or extremely toxic on already damaged cirrhotic livers. Prognosis prediction is based on clinicopathological parameters including tumour burden, proliferation markers, vascular invasion, liver function and overall health. Ongoing studies now incorporate biomarker candidates in clinical trials to test agents on patients who are most likely to benefit [13]. New single-gene prognostic marker candidates based on low-throughput technologies (e.g. polymerase chain reaction (PCR), immunohistochemistry (IHC)) appear in the literature almost on a daily basis. However, the generalizability of such markers remains questionable.

Vascular invasion of the portal or hepatic veins correlates strongly with HCC recurrence. Following liver transplantation, a 4.4-fold increased risk of HCC recurrence was present in patients with microvascular invasion [14]. At the same time, latent microscopic vascular invasion cannot be detected by pre-operative imaging [15]. Based on autopsy results, 40% of patients with tumours smaller than 5 cm already developed portal vein thrombi [16]. Therefore, identification of molecular changes that correlate with vascular invasion to ascertain the risk of HCC recurrence is a top priority to predict long-term outcome.

In the pursuit of robust prognostic genes, we conducted a meta-analysis by searching the literature for studies focusing on genes associated with OS. To cross-validate these biomarker candidates, we used a large available transcriptomic dataset representing roughly equally two ethnic groups (Asian and White/Caucasian). Prognostic potential for each gene was assessed in a univariate analysis within each ethnic cohort and the strongest markers were included in a multivariate regression. In addition, we also evaluated biomarkers associated with vascular invasion to validate their predictive potential in two independent datasets.

2. Material and methods

2.1. Identification of previously published biomarker candidates

A PubMed (http://www.pubmed.com/) search conducted in June 2017 using the keywords 'hepatocellular', 'carcinoma', 'overall' and 'survival' resulted in 22 999 hits and for terms 'hepatocellular', 'carcinoma', 'biomarker' and 'mRNA' in 1615 hits. The overlapping articles of the two sets established the database for cross-validation. All papers were included back to 1998. Only papers written in English were considered in the final biomarker selection. Only research articles were selected, excluding reviews. Biomarkers were retrieved only from studies involving patients undergoing surgical resection or liver transplantation. Only studies reporting molecular differences between tumour and healthy tissue were included in the final list, rejecting papers describing molecular differences between pre- or post-surgery serum or plasma levels.

As survival was frequently associated only with protein expression assessed by IHC or immunoblot, our analysis focusing only on studies reporting mRNA expression had to be extended to include protein-based assays as well. Studies not reporting follow-up data were excluded, same as for papers reporting non-significant associations with survival. Of note, some of the articles assessed the simultaneous co-expression of two to five genes, labelled as composed-biomarkers.

2.2. Transcriptomic dataset to validate prognostic biomarkers

The Gene Expression Omnibus (GEO), the European Genome-phenome Archive (EGA) and The Cancer Genome Atlas (TCGA) repositories were screened for datasets with available survival data and at least 30 patients. Accessible transcriptomic data of 371 patients diagnosed with HCC were obtained from the TCGA (https://cancergenome.nih.gov/), with OS available for 364 patients [17]. Most patients were diagnosed at stage I and 67% were male. Patients with Asian and White/Caucasian ethnicity made up the majority,

Table 2. Characterization of patient populations obtained from three independent transcriptomic datasets used in this study to validate biomarkers of overall survival and vascular invasion.

cohort	RNA-seq	Illumina	Affymetrix
source	TCGA	GSE20017	GSE9843
platform	Illumina HiSeq 2000	Illumina HumanRef-8 WG-DASL v. 3.0	Affymetrix HGU133 Plus 2.0 Array
total *n*	371	135	91
gender			
males	250	102	54
females	121	33	27
NA	—	—	10
stage			
stage I	171	—	9
stage II	86	—	56
stage III	85	—	7
stage IV	5	—	8
NA	—	135	11
race			
White/Caucasian	184	102	72
Black or African-American	17	4	3
Asian	158	28	4
vascular invasion	—	40	45
death event	130	32	23

roughly in equal proportions. There was no significant difference in survival when comparing Asian and White/Caucasian patients ($p = 0.32$). Detailed description of the sample population is included in table 2.

2.3. Biomarker candidates associated with vascular invasion

After the initial PubMed search for 'HCC' + 'biomarkers', the list of genes was further narrowed by keywords 'vascular invasion' and 'PVTT' (as for portal vein tumour thrombosis). Papers written in non-English and reviews were omitted from the final biomarker selection. The final list consists of studies reporting significant associations between mRNA/protein expression and vascular invasion/PVTT.

2.4. Transcriptomic datasets to validate predictive biomarkers

The GEO database was searched for keywords 'hepatocellular' + 'carcinoma' + 'vascular' + 'invasion' in May 2017. The search resulted in 18 hits in humans. Of these, 13 remained after filtering for expression profiling by the array. Of these, only two datasets (GSE9843 and GSE20017) had at least 50 patients. GSE9843 contained fresh frozen tissue samples from 91 patients, including 45 with and 34 without vascular invasion [7]. Of all patients, 91% were of White/Caucasian origin, 65.8% were male and 65 patients had published stage (table 2). In GSE20017, genome-wide gene expression profiling from formalin-fixed paraffin-embedded tissues of 135 patients was published. Of all patients, 75.6% were males, 102 of White/Caucasian origin and 40 patients were diagnosed with vascular invasion (table 2). The preprocessed and normalized gene expression data were used in both datasets.

2.5. Statistical analyses

A unique HUGO identifier was assigned to each biomarker candidate. Differences in OS were tested by Cox proportional hazards regression. Analyses including Kaplan–Meier survival plots, hazard rates with 95% confidence interval and log-rank *p*-values were calculated and plotted in R using Bioconductor

packages separately for the Asian and White/Caucasian ethnic groups and also for the pooled dataset. We considered a p-value of 0.05 as significant. The p-values from univariate analysis were corrected for multiple testing by computing the false discovery rate (FDR) as described previously [18]. We made a multivariate analysis to compare significant genes to available clinical variables. The expression level of genes was compared with a Mann–Whitney U-test between patients with or without vascular invasion independently in both datasets.

3. Results

3.1. Identification of previously described HCC biomarker candidates

The PubMed search for HCC prognostic biomarkers resulted in 533 hits, of which 513 were written in English, and seven papers were reviews. Evaluation of the remaining 506 papers resulted in 355 relevant articles describing differences between mRNA and/or protein expression and survival. For some genes, multiple references were available, decreasing the number of individual biomarkers to 318, from which overexpression of 194 genes was associated with poor OS. Only 17 papers relied on samples from White/Caucasian subjects describing 21 separate biomarkers. All the other studies involved patients from Asia, mainly from China but also from Korea, Taiwan and Japan.

Molecular association with survival relied frequently on protein expression. In fact, 57% of studies compared survival between groups with low versus high protein expression as measured by IHC, 38% used mRNA expression and the remaining 5% of studies exploited previously available datasets.

Thirteen papers explored the simultaneous expression of two to five genes, thus the cross-validation was conducted on 305 unique genes and 13 composed-biomarkers. The entire gene list with the corresponding manuscript identifier PMIDs is included in electronic supplementary material, table S1.

3.2. Prognostic biomarker candidates in Asian patients

Out of the 318 candidate markers, 180 were associated with OS at $p < 0.05$ (electronic supplementary material, table S2). In the pursuit of the strongest biomarkers, a rigorous correction for multiple testing conducted implementing a 0.1% FDR cut-off resulted in 40 significant biomarkers, consisting of 39 unique genes and a single composed-biomarker. High expression of eight genes was associated with improved survival, and expression of 32 biomarkers was linked to poor OS (table 3). Survival plots for the two best performing genes are illustrated in figure 1a,b.

To assess the robustness of our meta-analysis, we calculated the hazard rate for each of the 20 501 genes with available expression data to compute their association with OS. Importantly, of the genes identified by the literature search, the ones that remained significant after our validation were located among the strongest markers of survival (figure 2a,b).

3.3. Gene ontology analysis of prognostic biomarker candidates

The 40 potential prognostic biomarkers of the Asian cohort were subjected to gene enrichment analysis by Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources 6.8 to gauge the biological meaning of functionally related gene groups [59]. Gene sets related to various facets of cell cycle and cell division, such as sister chromatid cohesion, mitotic nuclear division and cyclin degradation were enriched significantly among markers of the Asian cohort, grouped by relevant function in electronic supplementary material, table S3.

3.4. Prognostic biomarker candidates in White/Caucasian patients

Out of 318 biomarkers, 128 were associated with OS at $p < 0.05$ in White/Caucasian subjects (electronic supplementary material, table S2). Out of 21 biomarkers that were appraised originally at White/Caucasian patients only, 10 remained significant at $p < 0.05$: *THOC5* ($p = 0.00018$), *PLAT* ($p = 0.0012$), *HMGA1* ($p = 0.0025$), *PRKDC* ($p = 0.0096$), *SLC22A1* ($p = 0.013$), *CDC20* ($p = 0.016$), the composed-biomarker *CD274-CXCL12* ($p = 0.019$), *CXCL12* ($p = 0.027$), *RRM2* ($p = 0.045$) and *ELAVL1* ($p = 0.049$). However, no biomarker candidates remained significant at 0.1% FDR. After a more lenient correction for multiple testing (at 5% FDR), two genes retained significance. Low expression of *SOCS2* (HR = 0.37, 95% CI = 0.23–0.59, $p = 1.3 \times 10^{-5}$) while elevated expression of *NAT10* (HR = 3.11, 95% CI = 1.76–5.52, $p = 4.4 \times 10^{-5}$) was coupled with the poor OS (figure 1c,d).

Table 3. List of 40 significant biomarker candidates associated with overall survival at 0.1% FDR in the Asian cohort. Cox univariate regression analysis. HR, hazard rate; CI, confidence interval.

symbol	gene name	overall survival HR (95% CI), p	worse prognosis (expression low – high)	references
BIRC5	baculoviral IAP repeat containing 5	HR = 7.64 (3.66 – 15.94) $p = 1.9 \times 10^{-10}$	high	[19 – 21]
CDC20	cell division cycle 20	HR = 5.53 (2.95 – 10.38) $p = 2.5 \times 10^{-9}$	high	[22]
PLK1	Polo-like kinase 1	HR = 6.02 (3.08 – 11.78) $p = 3 \times 10^{-9}$	high	[23]
ALDH2	aldehyde dehydrogenase-2	HR = 0.19 (0.1 – 0.35) $p = 4.2 \times 10^{-9}$	low	[24]
CCNB1	cyclin B1	HR = 7.09 (3.29 – 15.29) $p = 6.3 \times 10^{-9}$	high	[25]
FOXK2	forkhead box K2	HR = 6.83 (3.16 – 14.73) $p = 1.5 \times 10^{-8}$	high	[26]
KIF18A	kinesin family member 18A	HR = 4.9 (2.66 – 9.01) $p = 1.9 \times 10^{-8}$	high	[27]
BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B	HR = 4.85 (2.64 – 8.92) $p = 2.4 \times 10^{-8}$	high	[28]
MELK	maternal embryonic leucine zipper kinase	HR = 5.64 (2.84 – 11.22) $p = 3.1 \times 10^{-8}$	high	[29]
KIAA1524	cancerous inhibitor of protein phosphatase 2A	HR = 4.71 (2.57 – 8.63) $p = 4.0 \times 10^{-8}$	high	[30]
PTTG1	pituitary tumour-transforming 1	HR = 7.67 (3.24 – 18.2) $p = 5.2 \times 10^{-8}$	high	[31]
CCNF	cyclin F	HR = 4.72 (2.54 – 8.76) $p = 7.1 \times 10^{-8}$	high	[32]
PKM2	pyruvate kinase M2	HR = 4.48 (2.45 – 8.18) $p = 1 \times 10^{-7}$	high	[33,34]
DEPDC1	DEP domain containing 1	HR = 4.74 (2.47 – 9.1) $p = 2.6 \times 10^{-7}$	high	[35]
STMN1 and SPP1	stathmin 1 and secreted phosphoprotein 1 (osteopontin)	HR = 4.31 (2.34 – 7.92) $p = 3.2 \times 10^{-7}$	high	[36]
ANGPT2	angiopoietin 2	HR = 4.25 (2.32 – 7.76) $p = 3.3 \times 10^{-7}$	high	[37]
EZH2	enhancer of zeste 2 polycomb repressive complex 2 subunit	HR = 5.51 (2.64 – 11.51) $p = 3.5 \times 10^{-7}$	high	[38]
STMN1	stathmin 1	HR = 4.97 (2.5 – 9.88) $p = 4.2 \times 10^{-7}$	high	[39]
CDC25A	cell division cycle 25A	HR = 4.22 (2.28 – 7.8) $p = 6.7 \times 10^{-7}$	high	[40]

(*Continued.*)

Table 3. (Continued.)

| symbol	gene name	overall survival HR (95% CI), p	worse prognosis (expression low – high)	references
SLC22A1	solute carrier family 22 member 1	HR = 0.24 (0.13 – 0.45) $p = 8.4 \times 10^{-7}$	low	[41]
AURKA	aurora kinase A	HR = 4.5 (2.33 – 8.66) $p = 9.6 \times 10^{-7}$	high	[42]
SEC62	SEC62 homologue, preprotein translocation factor	HR = 0.25 (0.14 – 0.46) $p = 1.1 \times 10^{-6}$	low	[25]
FOXM1	forkhead box M1	HR = 4.5 (2.31 – 8.78) $p = 1.4 \times 10^{-6}$	high	[43]
ADH4	alcohol dehydrogenase 4 (Class II), Pi polypeptide	HR = 0.26 (0.14 – 0.47) $p = 1.8 \times 10^{-6}$	low	[44]
MKI67	marker of proliferation Ki-67	HR = 3.89 (2.13 – 7.12) $p = 2.2 \times 10^{-6}$	high	[22,45]
MAD2L1	MAD2 mitotic arrest deficient-like 1 (yeast)	HR = 9.92 (3.06 – 32.14) $p = 2.6 \times 10^{-6}$	high	[46]
CDK4	cyclin dependent kinase 4	HR = 6.8 (2.68 – 17.28) $p = 3.1 \times 10^{-6}$	high	[47]
CKAP2	cytoskeleton associated protein 2	HR = 3.64 (2.01 – 6.61) $p = 5.5 \times 10^{-6}$	high	[48]
IQGAP2	IQ motif containing GTPase activating protein 2	HR = 0.2 (0.09 – 0.44) $p = 7.3 \times 10^{-6}$	low	[49]
SOCS2	suppressor of cytokine signalling 2	HR = 0.26 (0.13 – 0.49) $p = 7.7 \times 10^{-6}$	low	[50]
E2F1	E2F transcription factor 1	HR = 8.66 (2.68 – 28.02) $p = 1.4 \times 10^{-5}$	high	[39]
RRM2	ribonucleotide reductase regulatory subunit M2	HR = 3.7 (1.96 – 6.98) $p = 1.6 \times 10^{-5}$	high	[22]
HMGA1	high mobility group AT-Hook 1	HR = 3.38 (1.86 – 6.13) $p = 2.1 \times 10^{-5}$	high	[51]
E2F3	E2F transcription factor 3	HR = 4.1 (2.02 – 8.31) $p = 2.3 \times 10^{-5}$	high	[52]
KIT	KIT proto-oncogene receptor tyrosine kinase	HR = 3.34 (1.84 – 6.05) $p = 2.5 \times 10^{-5}$	high	[53]
CKS1B	CDC28 protein kinase regulatory subunit 1B	HR = 4.06 (2 – 8.23) $p = 2.6 \times 10^{-5}$	high	[54]
FAM83D	family with sequence similarity 83 member D	HR = 3.58 (1.89 – 6.78) $p = 2.8 \times 10^{-5}$	high	[55]
CENPH	centromere protein H	HR = 4.05 (1.99 – 8.21) $p = 2.8 \times 10^{-5}$	high	[56]
PTEN	phosphatase and tensin homologue	HR = 0.3 (0.16 – 0.55) $p = 3.1 \times 10^{-5}$	low	[57]
HLX	H2.0 like homeobox	HR = 0.26 (0.13 – 0.52) $p = 3.4 \times 10^{-5}$	low	[58]
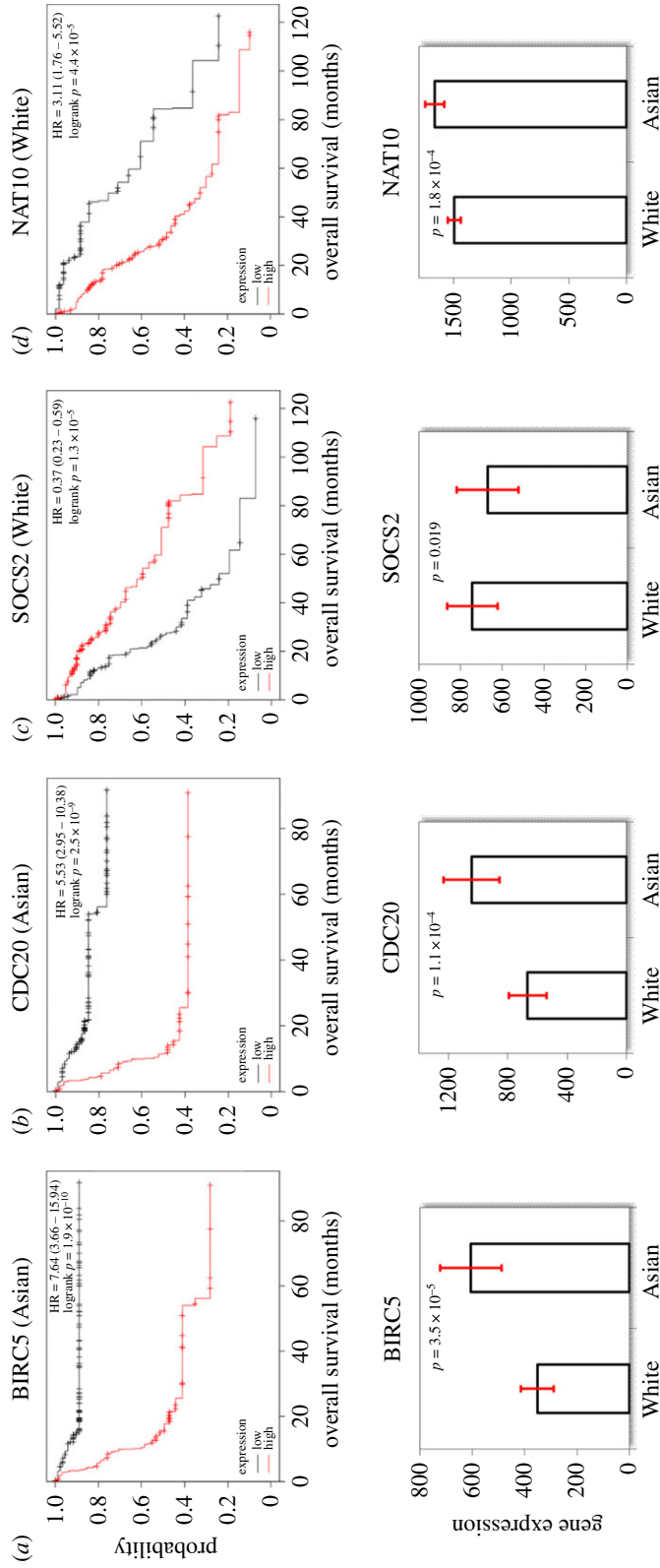
Figure 1. Survival for the two best performing genes in the Asian (*a*,*b*) and in the White/Caucasian cohorts (*c*,*d*). The expression of all four genes was significantly different between the Asian and White/Caucasian cohorts (bottom subfigures). Red bars: 95% confidence intervals.
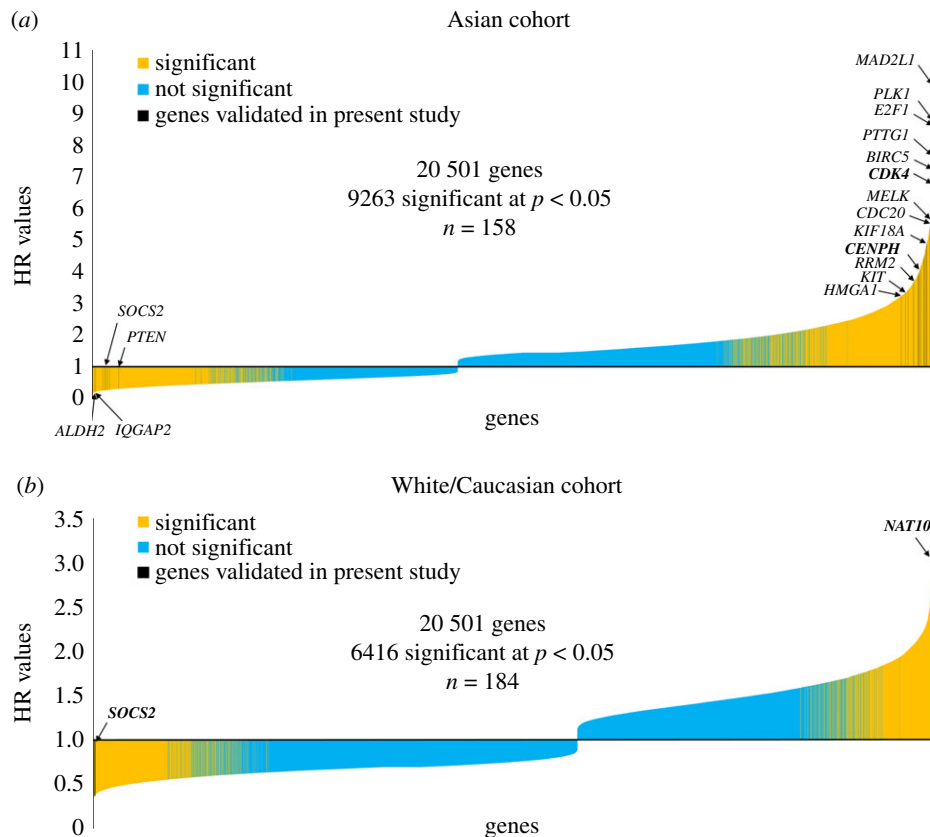
Figure 2. Ranked hazard rates across all 20 501 genes measured by RNA-seq in the Asian (*a*) and in the White/Caucasian cohort (*b*). Genes significantly associated with overall survival are yellow-coloured. The 39 strongest single-gene biomarker candidates validated in the Asian cohort are indicated by black lines including names for selected genes—notably, all positioned among the strongest markers of overall survival. Genes significant after multivariate analysis in the Asian cohort are indicated in bold face, as well as the two biomarker candidates significant at 5% FDR in the White/Caucasian cohort.

3.5. Overlap in significant biomarker candidates between Asian and White/Caucasian cohorts

Out of the 318 biomarker candidates, 226 were significant either in the Asian or in the White/Caucasian cohort at $p < 0.05$. Out of them, 98 and 46 biomarkers were unique to Asian and White/Caucasian patients, respectively, and 82 biomarker candidates were shared by both ethnic groups (electronic supplementary material, table S2). Among the shared biomarkers, 72 were originally described in Asian and 10 in White/Caucasian subjects. Although *NAT10* was not among the 40 significant biomarkers associated with OS in the Asian cohort after setting the threshold at 0.1% FDR, *SOCS2* was among the strongest biomarkers in both ethnic groups.

Significant biomarker candidates from univariate analysis of each ethic group were subjected to gene list analysis by the PANTHER (Protein ANalysis THrough Evolutionary Relationships, http://pantherdb.org) gene classification system [60]. Using the pathway classification function, the total number of pathway hits were 132 and 67 in the Asian and White cohorts, respectively. Pathways represented by at least three genes in the Asian and by at least two genes in the White/Caucasian cohort are illustrated in figure 3. Genes participating in FGF-, IGF-signalling, hypoxia–response, CCKR-, chemokine- and cytokine-signalling pathways were overrepresented in the Asian population compared to Caucasians, while genes participating in Wnt-signalling, Alzheimer disease-, presenilin-, PDGF-, cadherin-, JAK/STAT- and GnRHR-signalling pathways were more heavily represented in the Caucasian cohort.

3.6. Prognostic biomarker candidates in the pooled dataset

OS did not differ among ethnic groups ($p = 0.32$). After merging patient data with different ethnicity, 178 genes remained associated with OS at $p < 0.05$. Only 40 of these markers reached significance for OS in the analysis at 0.1% FDR, incorporating a single composed-biomarker of two genes and 39 unique genes.
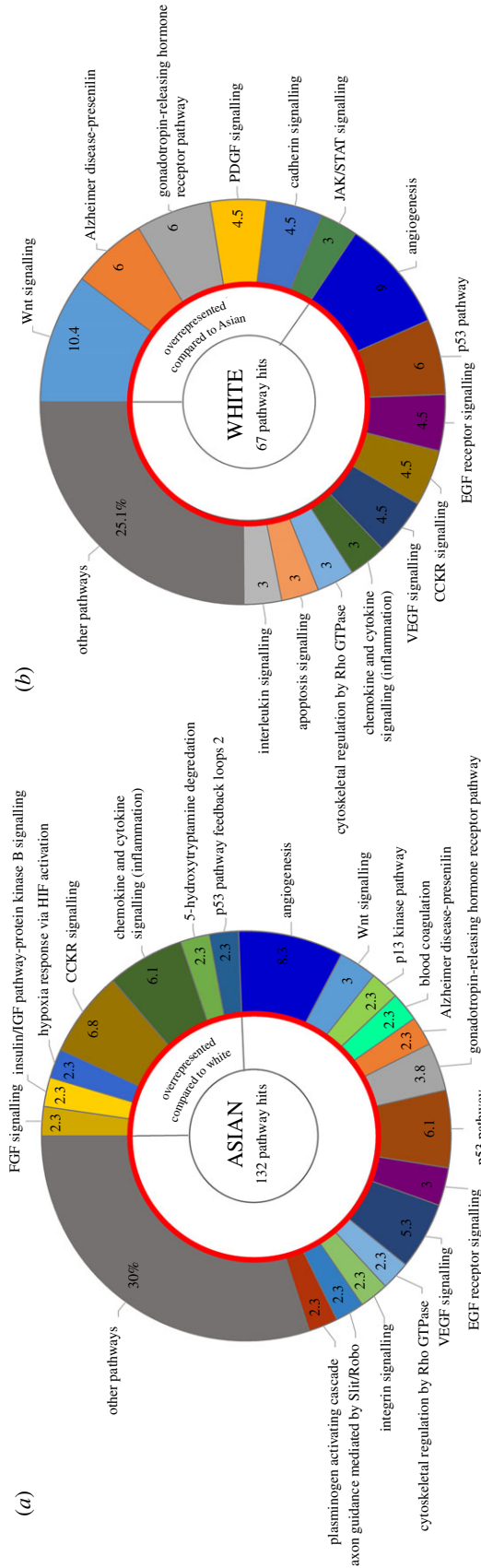
Figure 3. Comparison of gene ontology pathways in Asian (*a*) and White/Caucasian (*b*) patients based on PANTHER gene classification system. All genes significant in the univariate analysis are included. The total number of pathway hits (132 and 67 in the Asian and White cohorts, respectively) represent 100%. Emphasized are those functions that are overrepresented by at least 2% in a given ethnic group compared to the other cohort.

High expression of 17 genes correlated with a better outcome, while the high expression of 23 biomarker candidates indicated poor OS (electronic supplementary material, table S4).

3.7. Multivariate analysis

In the multivariate analysis, we included parameters which were available for the majority of patients. We had to exclude other clinically relevant features like differentiation, size, cirrhosis and alpha-fetoprotein (AFP) levels because these were only published for a few patients—a robust multivariate analysis can only take account of patients where all data are simultaneously available. In Asian samples, multivariate analysis for OS included stage, sex and expression of 39 unique single-gene biomarkers significant at 0.1% FDR. OS was strongly associated with stage ($p = 0.0018$). Two genes remained prognostic for OS: *CENPH* ($p = 0.0038$) and *CDK4* ($p = 0.038$). Marginal significance was reached by *CDC20* ($p = 0.053$). High expression of all three genes (*CENPH*, *CDK4* and *CDC20*) predicted poor OS.

In the Caucasian cohort, only two genes remained significant after correction for multiple hypothesis testing (5% FDR); therefore, multivariate analysis included stage, sex and expression of these two genes, *SOCS2* and *NAT10*. OS was associated with the expression of both *SOCS2* ($p = 0.046$) and *NAT10* ($p = 0.031$), but stage and sex were not prognostic.

In the pooled dataset, multivariate analysis for OS including stage, sex and expression of 39 unique, single-gene biomarkers resulted in a significant association between OS and stage ($p = 2.06 \times 10^{-5}$) and the expression of eight genes: *SPP1* ($p = 0.0002$), *PKM2* ($p = 0.0008$), *EZH2* ($p = 0.0017$), *KIF18A* ($p = 0.0061$), *DEPDC1* ($p = 0.0092$), *CCNF* ($p = 0.019$), *SLC2A1* ($p = 0.036$) and *WASF2* ($p = 0.048$). Increased expression of all eight genes was associated with poor survival outcome.

3.8. Biomarkers predictive of vascular invasion

The keywords, vascular 'invasion' and 'PVTT' narrowed the list of previously filtered 506 papers to 66 and 8 hits, respectively, scaling down our original list of biomarker candidates to 52 unique genes (electronic supplementary material, table S5).

In GSE9843, information about the presence of vascular invasion was available for 79 patients, out of which 45 patients presented with vascular invasion. Out of the 52 biomarker candidates, the expression of 9 genes was significantly different between patients with or without vascular invasion at $p < 0.05$ (*GPC3*, *KIF18A*, *NDRG1*, *PYGO2*, *RORA*, *STMN1*, *VEGFB*, *VIL1* and *XAF1*) (table 4) and an additional 4 genes reached marginal significance (*CADM2*, *FAM83D*, *NAA10* and *PEBP1*). The expression of *GPC3*, *KIF18A*, *NDRG1*, *PYGO2*, *STMN1*, *VEGFB* and *VIL1* was increased and the expression of *RORA* and *XAF1* was decreased in patients with vascular invasion (figure 4*a*).

In GSE20017, 40 out of 135 patients were diagnosed with vascular invasion. Expression of 10 genes (*APOA1*, *ENG*, *FABP1*, *ID2*, *KIF18A*, *NEDD9*, *NODAL*, *PDGFRA*, *PTP4A3*, *STAT4*) was significantly different between patients with and without vascular invasion at $p < 0.05$ (table 4), and expression difference of *ECM1* and *RPS19BP1* reached marginal significance. The expression of *APOA1*, *ENG*, *FABP1*, *ID2*, *NODAL*, *PDGFRA* and *STAT4* was decreased and the expression of *KIF18A*, *NEDD9* and *PTP4A3* was increased in patients with vascular invasion (figure 4*b*).

KIF18A was the only common gene across the two datasets significantly different correlated to vascular invasion (table 4) and its expression was significantly higher in patients with vascular invasion in both datasets (figure 4*c*).

4. Discussion

The lack of targeted therapies in HCC coupled with escalating incidence called forth a paradigm shift. After a decade of phase III failures, current trials started to incorporate molecular markers, and targeted treatments are offered to preselected patients [13]. To accommodate the needs, abundant hypothesis-driven prognostic biomarkers are published in the literature, mostly assessing the transcriptome or proteome by low-throughput technologies (qPCR, IHC, immunoblot). We aimed to cross-validate these published markers to gauge their clinical generalizability. We analysed 318 markers disclosed as related to HCC prognosis in studies published back to 1998. The strongest candidates were also subjected to a multivariate regression. Vascular invasion is one of the strongest clinicopathological features predicting long-term outcome after resection or transplantation. Fifty-two genes associated with vascular invasion were validated in two independent datasets.

Table 4. Genes associated with the presence of vascular invasion based on the Affymetrix and Illumina datasets. Genes were differently expressed between patients with or without vascular invasion at $p < 0.05$. KIF18A is the only gene significant in both datasets.

symbol	gene name	Affymetrix p-value	Illumina p-value
KIF18A	kinesin family member 18A	0.003	0.025
GPC3	glypican-3	0.022	n.s.
NDRG1	N-Myc downstream regulated 1	0.041	n.s.
PYGO2	pygopus family PHD finger 2	0.03	n.s.
RORA	RAR-related orphan receptor A	0.005	n.s.
STMN1	stathmin 1	0.022	n.s.
VEGFB	vascular endothelial growth factor B	0.014	n.s.
VIL1	villin-1	0.022	n.s.
XAF1	XIAP associated factor 1	0.039	n.s.
APOA1	apolipoprotein A1	n.s.	0.0002
ENG	endoglin	n.s.	0.028
FABP1	fatty acid binding protein 1	n.s.	0.043
ID2	inhibitor of DNA binding 2, HLH protein	n.s.	0.01
NODAL	nodal growth differentiation factor	n.s.	0.029
PDGFRA	platelet-derived growth factor receptor alpha	n.s.	0.004
STAT4	signal transducer and activator of transcription 4	n.s.	0.001
NEDD9	neural precursor cell expressed, developmentally downregulated 9	n.s.	0.025
PTP4A3	protein tyrosine phosphatase type IVA, member 3	n.s.	0.048

Our results highlight the importance of independent cross-validations, as from 318 markers only 180 and 128 genes remained prognostic in the Asian and White/Caucasian cohorts, respectively. Only 40 markers retained significance after rigorous correction for multiple comparisons in the Asian cohort, and none remained significant in the White/Caucasian cohort. This extremely high attrition rate is partially the consequence of small sample sizes in the original studies. Furthermore, in many instances, the transparent and complete reporting allowing the evaluation of the soundness of the study still did not reach the required standards set by the REMARK guidelines [2]. We have to note that survival was frequently assessed as a function of protein expression not inevitably congruent with the transcriptome.

Identifying major components of vascular invasion is vital for a successful therapeutic intervention. Out of the 52 genes described in the literature, 9 and 10 biomarkers were significantly differentially expressed between patients with or without vascular invasion in GSE9843 and GSE20017, respectively. However, of the 18 significant genes, KIF18A was the solitary common gene across both datasets. KIF18A also appeared among the 40 strongest biomarkers of survival in the Asian cohort. KIF18A mediates transport of organelles, proteins and plays a role in microtubule motor activity and mitotic chromosome alignment during cell division [61]. KIF18A has also been associated with metastasis in solid tumours (e.g. breast cancer [62]). Specific kinesin motor proteins and molecules participating in cell cycle can be potentially targeted [63].

In the presence of vascular invasion, most biomarkers were overexpressed in GSE9843, but downregulated in GSE20017. One additional strong prognostic biomarker, STMN1, was also overexpressed in patients with vascular invasion; however, only in one of the datasets. The lack of overlap highlights the heterogeneity of HCC samples across different datasets, questions the generalizability of the studied biomarkers and emphasizes the importance of external validation.

SPP1 (osteopontin) was the strongest biomarker candidate associated with survival in the pooled dataset; moreover, SPP1 expression was significantly associated with worse outcome both in the Asian and White/Caucasian datasets. SPP1 is a multitasking, highly phosphorylated extracellular matrix protein affecting ECM degradation, cell motility and adhesion, cytoskeletal rearrangement, mitosis and inflammation [64,65]. SPP1 has been previously recognized as a potential marker of early
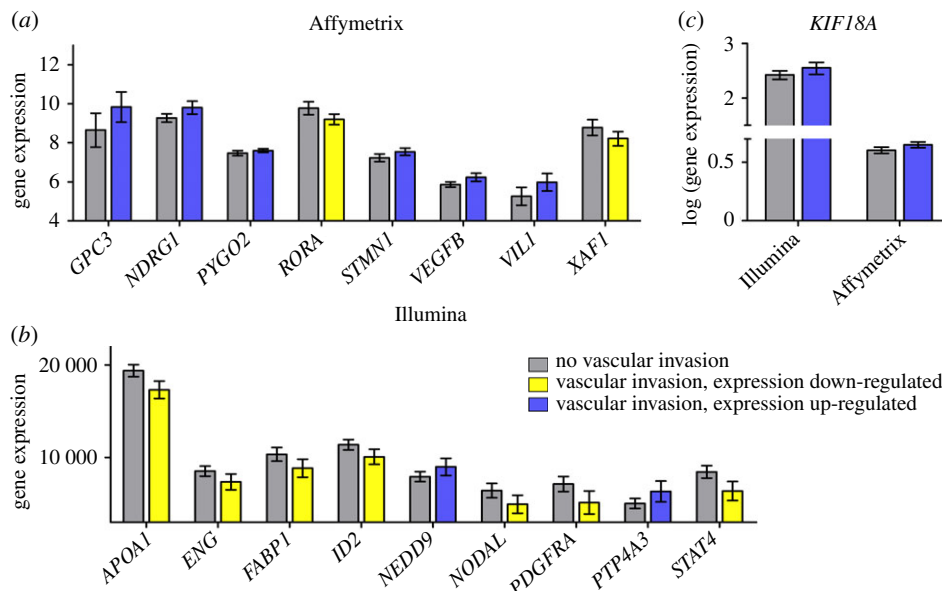
Figure 4. Genes differently expressed between patients with or without vascular invasion at $p < 0.05$ in the Affymetrix (*a*) and Illumina datasets (*b*). In Affymetrix, *RORA* and *XAF1* were downregulated (yellow) with the remaining genes upregulated in the presence of vascular invasion (blue). In Illumina, *NEDD9* and *PTP4A3* (blue) along with *KIF18A* showed increased expression in patients with vascular invasion. *KIF18A* was the singular common gene upregulated in patients with vascular invasion in both datasets (*c*). Error bars: 95% confidence intervals.

recurrence and poor prognosis and as a leading metastasis-related gene in HCC [66,67]. A meta-analysis including seven studies confirmed that plasma *SPP1* elevation and AFP-based results have comparable diagnostic performance [68,69], although elevated *SPP1* may be linked to other malignancies, therefore should be combined with other HCC-specific biomarkers [70]. Numerous *SPP1* products are generated by genetic polymorphisms, alternative spicing and posttranslational modifications [65]. Various preclinical therapeutic approaches are being evaluated to target *SPP1*, although when interfering with its normal function severe adverse effects may develop [71].

In the Asian cohort, the strongest biomarkers of survival were particularly enriched by genes controlling multiple aspects of cell division, including sister chromatid cohesion and positioning (*CENPH, KIF18A*), mitotic checkpoint (*BUB1B, MAD2L1*), cell cycle transitions (*CCNB1, PLK1, FOXM1, BIRC5, AURKA, CDC25A, MELK, CCNF*), G1-phase progression (*CDK4, E2F1*) and cyclin degradation (*CDC20, PTTG1*). Various efforts classified HCC into non-proliferative and proliferative genotypes, with worse outcome in the latter [13]. Our results are congruent with previous findings, where the G3 tumour type overexpressing genes controlling cell cycle was associated with the most severe prognosis [72], and G3 type was confirmed as an independent predictor of recurrence [73].

BIRC5 (survivin) was the strongest predictor of poor survival based on the univariate analysis in the Asian cohort (figure 1). *BIRC5* is the smallest member of the inhibitor of apoptosis protein (IAP) family, a multitasking protein that regulates mitosis, suppresses apoptosis and affects proliferation, angiogenesis and cellular stress response [74–76]. *BIRC5* is almost entirely absent in adult tissue, but its upregulation is linked to worse survival in numerous solid tumours [77–79]. A systematic meta-analysis confirmed the prognostic role of *BIRC5* gene in HCC across 14 different studies involving 890 patients [80]. The multiple and redundant regulation of *BIRC5* poses a difficulty for inhibitor development and requires simultaneous targeting of multiple networks [81,82]. *BIRC5* deficiency interferes with mitosis, increases TNFα and leads to senescence. Promising tumour regression was reported after synergistic effects of a mitotic inhibitor acting on *PLK1* (BI2536) to induce mitotic arrest, and a proapoptotic agent (LCL161) to sensitize HCC cells to TNFα-triggered cell death in a murine model of HCC [83].

Two genes remained independent prognostic factors after multivariate analysis in the Asian cohort: *CENPH* and *CDK4*, both linked to poor prognosis. Centromeres and their associated kinetochores are required for proper spindle attachment, separation of sister chromatids and regulation of mitotic checkpoint activity [84]. The product of *CENPH* is a fundamental component of the inner kinetochore–nucleosome associated complex [85]. Misregulation of centromere and kinetochore protein levels is linked to missegregations, aneuploidy and genome rearrangement [86,87], increasing chromosomal

instability (CIN) and tumourigenesis. Accumulating evidence suggests that *CENPH* upregulation has been linked to aneuploidy, high proliferative ability and poor prognosis in various solid tumours [88–90].

CDK4 acts on the retinoblastoma (RB) signalling pathway, promoting progression from G1 to S phase during the cell cycle. A selective CDK4/6 inhibitor, palbociclib, became recently approved as first-line treatment for oestrogen-positive, HER2-negative breast cancer patients. The majority of HCCs harbour an intact *RB1* gene [91], rendering them potentially sensitive to palbociclib. In HCC, palbociclib suppressed *in vitro* cell proliferation, restricted *in vivo* tumour growth and increased survival with enhanced effects when combined with sorafenib [92].

In the White cohort only genes remained significant after *p*-value correction, possibly as a consequence of most articles focusing at Asian patients; in fact, only 6% of papers analysed samples from White/Caucasian subjects. Of *SOCS2* and *NAT10,* only the latter offers a potential target for future therapies. *NAT10* is involved in histone-acetylation and regulates telomerase activity through TERT-promoter [93]. It is localized mainly in the cell membrane, particularly at the invasive leading edge of tumours, increasing motility and invasive potential that correlates with nuclear β-catenin accumulation [94,95]. *NAT10* also activates p53 in response to DNA damage by acetylation [96]. *NAT10* is the target of a small molecule compound called 'remodelin' to treat laminopathies and premature ageing syndromes [97], not yet tested in the context of HCC.

Among our strongest biomarkers of survival, several genes have been repeatedly identified in the process of oncogenic transformation, including the development of LC [98]. Our gene ontology analyses revealed ethnic group-specific signalling pathway activations (figure 3). The differences are probably related to the distinct aetiology of the disease, as viral infections underlie most HCCs in Asia, while in western countries, chronic liver disease related to alcohol consumption and non-alcoholic steatohepatitis is responsible for HCC [6].

Our narrowed list of biomarkers may provide optimal targets to tackle pivotal signalling pathways of HCC including the Wnt/β-catenin, TP53, RAS, JAK–STAT, MAPK, the retinoblastoma protein pRb1, oxidative stress and chromatin remodelling pathways [13]. *PKM2* is the rate-limiting catalytic activator of glycolytic metabolic pathways and also exerts a non-glycolytic function as a transcriptional activator, possibly affecting numerous signalling pathways, including Wnt/β-catenin, JAK-STAT and AKT signalling [99]. P53 inactivation is abundant in HCC even in the absence of *TP53* mutations, leading to overexpression of genes normally repressed by p53 such as *EZH2,* a histone-methyltransferase that epigenetically regulates stem cell maintenance, and *CCNB1, E2F1* and *FOXM1,* all present in our list of strongest prognostic markers [17]. Dosage-dependent copy number gains of *E2F1* inactivate the retinoblastoma tumour suppressor pathway, driving HCC development in mice [100].

Several of our strongest prognostic biomarker candidates have also been identified with a functionally significant role in drug response. For example, the tyrosine kinase receptor KIT is a target of the kinase inhibitor sorafenib [101]. Moreover, sorafenib decreases protein levels of survivin, the product of the anti-apoptotic *BIRC5* [102] and expression of numerous genes implicated in cell cycle, DNA replication and cell proliferation (e.g. *CDC20, CDC25A*) [103]. Our list of significant prognostic biomarkers might help to select candidates for future drug development.

Altogether, here we aimed to deliver authentication for gene- and protein expression-based HCC biomarkers by assessing the generalizability of previous findings. Integration of copy number variations, somatic mutations, methylation patterns, proteomics and mRNA expression might provide more throughout the identification of factors related to HCC [17]. Additional RNAs complementing mRNA started to emerge as potential biomarkers and therapeutic targets. A recent paper identified 223 miRNAs associated with HCC and validated their significance in OS [104].

5. Conclusion

The low reproducibility of the published prognostic biomarkers uncovered in our analysis draws attention to the need for more rigorous research practices in this heterogeneous neoplastic disease with a highly variable aetiology. In summary, we provide a unified pipeline using already available transcriptomic datasets to enable the validation and ranking of future biomarker candidates.

15

References

1. Kern SE. 2012 Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. *Cancer Res.* **72**, 6097–6101. (doi:10.1158/0008-5472.CAN-12-3232)

2. Mallett S, Timmer A, Sauerbrei W, Altman DG. 2010 Reporting of prognostic studies of tumour markers: a review of published articles in relation to REMARK guidelines. *Br. J. Cancer* **102**, 173–180. (doi:10.1038/sj.bjc.6605462)

3. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. 2015 Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **136**, E359–E386. (doi:10.1002/ijc.29210)

4. Blachier M, Leleu H, Peck-Radosavljevic M, Valla DC, Roudot-Thoraval F. 2013 The burden of liver disease in Europe: a review of available epidemiological data. *J. Hepatol.* **58**, 593–608. (doi:10.1016/j.jhep.2012.12.005)

5. Ryerson AB et al. 2016 Annual report to the nation on the status of cancer, 1975–2012, featuring the increasing incidence of liver cancer. *Cancer* **122**, 1312–1337. (doi:10.1002/cncr.29936)

6. Yang JD, Roberts LR. 2010 Hepatocellular carcinoma: a global view. *Nat. Rev. Gastroenterol. Hepatol.* **7**, 448–458. (doi:10.1038/nrgastro.2010.100)

7. Minguez B et al. 2011 Gene-expression signature of vascular invasion in hepatocellular carcinoma. *J. Hepatol.* **55**, 1325–1331. (doi:10.1016/j.jhep.2011.02.034)

8. Llovet JM et al. 2008 Sorafenib in advanced hepatocellular carcinoma. *N. Engl. J. Med.* **359**, 378–390. (doi:10.1056/NEJMoa0708857)

9. Cheng AL et al. 2009 Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. *Lancet Oncol.* **10**, 25–34. (doi:10.1016/S1470-2045(08)70285-7)

10. Kudo M et al. 2018 Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial. *Lancet* **391**, 1163–1173. (doi:10.1016/S0140-6736(18)30207-1)

11. Bruix J et al. 2017 Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* **389**, 56–66. (doi:10.1016/S0140-6736(16)32453-9)

12. Schulze K, Nault JC, Villanueva A. 2016 Genetic profiling of hepatocellular carcinoma using next-generation sequencing. *J. Hepatol.* **65**, 1031–1042. (doi:10.1016/j.jhep.2016.05.035)

13. Llovet JM, Villanueva A, Lachenmayer A, Finn RS. 2015 Advances in targeted therapies for hepatocellular carcinoma in the genomic era. *Nat. Rev. Clin. Oncol.* **12**, 408–424. (doi:10.1038/nrclinonc.2015.103)

14. Iwatsuki S, Dvorchik I, Marsh JW, Madariaga JR, Carr B, Fung JJ, Starzl TE. 2000 Liver transplantation for hepatocellular carcinoma: a proposal of a prognostic scoring system. *J. Am. Coll. Surg.* **191**, 389–394. (doi:10.1016/S1072-7515(00)00688-8)

15. Zavaglia C et al. 2005 Predictors of long-term survival after liver transplantation for hepatocellular carcinoma. *Am. J. Gastroenterol.* **100**, 2708–2716. (doi:10.1111/j.1572-0241.2005.00289.x)

16. Yuki K, Hirohashi S, Sakamoto M, Kanai T, Shimosato Y. 1990 Growth and spread of hepatocellular carcinoma: a review of 240 consecutive autopsy cases. *Cancer* **66**, 2174–2179. (doi:10.1002/1097-0142(19901115)66:10<2174::AID-CNCR2820661022>3.0.CO;2-A)

17. Ally A et al. 2017 Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* **169**, 1327–1341. (doi:10.1016/j.cell.2017.05.046)

18. Gyorffy B, Gyorffy A, Tulassay Z. 2005 The problem of multiple testing and solutions for genome-wide studies. *Orv. Hetil.* **146**, 559–563.

19. Chau GY, Lee AF, Tsay SH, Ke YR, Kao HL, Wong FH, Tsou A-P, Chau Y-P. 2007 Clinicopathological significance of survivin expression in patients with hepatocellular carcinoma. *Histopathology* **51**, 204–218. (doi:10.1111/j.1365-2559.2007.02738.x)

20. Morinaga S et al. 2004 Expression of survivin mRNA associates with apoptosis, proliferation and histologically aggressive features in hepatocellular carcinoma. *Oncol. Rep.* **12**, 1189–1194. (doi:10.3892/or.12.6.1189)

21. Ikeguchi M, Ueta T, Yamane Y, Hirooka Y, Kaibara N. 2002 Inducible nitric oxide synthase and survivin messenger RNA expression in hepatocellular carcinoma. *Clin. Cancer Res.* **8**, 3131–3136.

22. Rebouissou S et al. 2017 Proliferation markers are associated with MET expression in hepatocellular carcinoma and predict tivantinib sensitivity in vitro. *Clin. Cancer Res.* **23**, 4364–4375. (doi:10.1158/1078-0432.CCR-16-3118)

23. He ZL, Zheng H, Lin H, Miao XY, Zhong DW. 2009 Overexpression of polo-like kinase1 predicts a poor prognosis in hepatocellular carcinoma patients. *World J. Gastroenterol.* **15**, 4177–4182. (doi:10.3748/wjg.15.4177)

24. Hou G et al. 2017 Aldehyde dehydrogenase-2 (ALDH2) opposes hepatocellular carcinoma progression by regulating AMP-activated protein kinase signaling in mice. *Hepatology* **65**, 1628–1644. (doi:10.1002/hep.29006)

25. Weng L, Du J, Zhou Q, Cheng B, Li J, Zhang D, Ling C. 2012 Identification of cyclin B1 and Sec62 as biomarkers for recurrence in patients with HBV-related hepatocellular carcinoma after surgical resection. *Mol. Cancer* **11**, 39. (doi:10.1186/1476-4598-11-39)

26. Lin MF, Yang YF, Peng ZP, Zhang MF, Liang JY, Chen W, Liu X-H, Zheng Y-L. 2017 FOXK2, regulated by miR-1271–5p, promotes cell growth and indicates unfavorable prognosis in hepatocellular carcinoma. *Int. J. Biochem. Cell Biol.* **88**, 155–161. (doi:10.1016/j.biocel.2017.05.019)

27. Liao W, Huang G, Liao Y, Yang J, Chen Q, Xiao S, Jin J, He S, Wang C. 2014 High KIF18A expression correlates with unfavorable prognosis in primary hepatocellular carcinoma. *Oncotarget* **5**, 10 271–10 279. (doi:10.18632/oncotarget.2082)

28. Liu AW, Cai J, Zhao XL, Xu AM, Fu HQ, Nian H, Zhang S-H. 2009 The clinicopathological significance of BUBR1 overexpression in hepatocellular carcinoma. *J. Clin. Pathol.* **62**, 1003–1008. (doi:10.1136/jcp.2009.066944)

29. Hiwatashi K et al. 2016 Expression of maternal embryonic leucine zipper kinase (MELK) correlates to malignant potentials in hepatocellular carcinoma. *Anticancer Res.* **36**, 5183–5188. (doi:10.21873/anticanres.11088)

30. He H, Wu G, Li W, Cao Y, Liu Y. 2012 CIP2A is highly expressed in hepatocellular carcinoma and predicts poor prognosis. *Diagn. Mol. Pathol.* **21**, 143–149. (doi:10.1097/PDM.0b013e318249fd8b)

31. Fujii T et al. 2006 Overexpression of pituitary tumor transforming gene 1 in HCC is associated with angiogenesis and poor prognosis. Hepatology 43, 1267–1275. (doi:10.1002/hep.21181)

32. Fu J, Qiu H, Cai M, Pan Y, Cao Y, Liu L, Yun J, Zhang CZ. 2013 Low cyclin F expression in hepatocellular carcinoma associates with poor differentiation and unfavorable prognosis. Cancer Sci. 104, 508–515. (doi:10.1111/cas.12100)

33. Hu W et al. 2015 Pyruvate kinase M2 prevents apoptosis via modulating Bim stability and associates with poor outcome in hepatocellular carcinoma. Oncotarget 6, 6570–6583. (doi:10.18632/oncotarget.3262)

34. Liu AM et al. 2014 miR-122 targets pyruvate kinase M2 and affects metabolism of hepatocellular carcinoma. PLoS ONE 9, e86872. (doi:10.1371/journal.pone.0086872)

35. Yuan SG, Liao WJ, Yang JJ, Huang GJ, Huang ZQ. 2014 DEP domain containing 1 is a novel diagnostic marker and prognostic predictor for hepatocellular carcinoma. Asian Pac. J. Cancer Prev. 15, 10917–10922. (doi:10.7314/APJCP.2014.15.24.10917)

36. Yuan RH, Jeng YM, Chen HL, Lai PL, Pan HW, Hsieh FJ, Lin C-Y, Lee P-H, Hsu H-C. 2006 Stathmin overexpression cooperates with p53 mutation and osteopontin overexpression, and is associated with tumour progression, early recurrence, and poor prognosis in hepatocellular carcinoma. J. Pathol. 209, 549–558. (doi:10.1002/path.2011)

37. Kuboki S et al. 2008 Angiopoietin-2 levels in the hepatic vein as a useful predictor of tumor invasiveness and prognosis in human hepatocellular carcinoma. J. Gastroenterol. Hepatol. 23, e157–e164. (doi:10.1111/j.1440-1746.2007.05175.x)

38. Hung SY, Lin HH, Yeh KT, Chang JG. 2014 Histone-modifying genes as biomarkers in hepatocellular carcinoma. Int. J. Clin. Exp. Pathol. 7, 2496–2507.

39. Chen YL et al. 2013 The E2F transcription factor 1 transactives stathmin 1 in hepatocellular carcinoma. Ann. Surg. Oncol. 20, 4041–4054. (doi:10.1245/s10434-012-2519-8)

40. Xu X et al. 2003 Overexpression of CDC25A phosphatase is associated with hypergrowth activity and poor prognosis of human hepatocellular carcinomas. Clin. Cancer Res. 9, 1764–1772.

41. Heise M et al. 2012 Downregulation of organic cation transporters OCT1 (SLC22A1) and OCT3 (SLC22A3) in human hepatocellular carcinoma and their prognostic significance. BMC Cancer 12, 109. (doi:10.1186/1471-2407-12-109)

42. Jeng YM, Peng SY, Lin CY, Hsu HC. 2004 Overexpression and amplification of Aurora-A in hepatocellular carcinoma. Clin. Cancer Res. 10, 2065–2071. (doi:10.1158/1078-0432.CCR-1057-03)

43. Sun H et al. 2011 FOXM1 expression predicts the prognosis in hepatocellular carcinoma patients after orthotopic liver transplantation combined with the Milan criteria. Cancer Lett. 306, 214–222. (doi:10.1016/j.canlet.2011.03.009)

44. Wei RR, Zhang MY, Rao HL, Pu HY, Zhang HZ, Wang HY. 2012 Identification of ADH4 as a novel and potential prognostic marker in hepatocellular carcinoma. Med. Oncol. 29, 2737–2743. (doi:10.1007/s12032-011-0126-3)

45. Kitamura K et al. 2011 Proliferative activity in hepatocellular carcinoma is closely correlated with glucose metabolism but not angiogenesis. J. Hepatol. 55, 846–857. (doi:10.1016/j.jhep.2011.01.038)

46. Zhang SH, Xu AM, Chen XF, Li DH, Sun MP, Wang YJ. 2008 Clinicopathologic significance of mitotic arrest defective protein 2 overexpression in hepatocellular carcinoma. Hum. Pathol. 39, 1827–1834. (doi:10.1016/j.humpath.2008.06.003)

47. Lu JW, Lin YM, Chang JG, Yeh KT, Chen RM, Tsai JJ, Su W-W, Hu R-M. 2013 Clinical implications of deregulated CDK4 and Cyclin D1 expression in patients with human hepatocellular carcinoma. Med. Oncol. 30, 379. (doi:10.1007/s12032-012-0379-5)

48. Hayashi T et al. 2014 Cytoskeleton-associated protein 2 is a potential predictive marker for risk of early and extensive recurrence of hepatocellular carcinoma after operative resection. Surgery 155, 114–123. (doi:10.1016/j.surg.2013.06.009)

49. Xia FD, Wang ZL, Chen HX, Huang Y, Li JD, Wang ZM, Li X-Y. 2014 Differential expression of IQGAP1/2 in hepatocellular carcinoma and its relationship with clinical outcomes. Asian Pac. J. Cancer Prev. 15, 4951–4956. (doi:10.7314/APJCP.2014.15.12.4951)

50. Cui M et al. 2016 The suppressor of cytokine signaling 2 (SOCS2) inhibits tumor metastasis in hepatocellular carcinoma. Tumour Biol. 37, 13521–13531. (doi:10.1007/s13277-016-5215-7)

51. Andreozzi M et al. 2016 HMGA1 expression in human hepatocellular carcinoma correlates with poor prognosis and promotes tumor growth and migration in in vitro models. Neoplasia 18, 724–731. (doi:10.1016/j.neo.2016.10.002)

52. Zeng X, Yin F, Liu X, Xu J, Xu Y, Huang J, Nan Y, Qiu X. 2014 Upregulation of E2F transcription factor 3 is associated with poor prognosis in hepatocellular carcinoma. Oncol. Rep. 31, 1139–1146. (doi:10.3892/or.2014.2968)

53. Chung CY, Yeh KT, Hsu NC, Chang JH, Lin JT, Horng HC, Chang C-S. 2005 Expression of c-kit protooncogene in human hepatocellular carcinoma. Cancer Lett. 217, 231–236. (doi:10.1016/j.canlet.2004.06.045)

54. Huang CW et al. 2010 CKS1B overexpression implicates clinical aggressiveness of hepatocellular carcinomas but not p27(Kip1) protein turnover: an independent prognosticator with potential p27 (Kip1)-independent oncogenic attributes? Ann. Surg. Oncol. 17, 907–922. (doi:10.1245/s10434-009-0779-8)

55. Liao W et al. 2015 Upregulation of FAM83D affects the proliferation and invasion of hepatocellular carcinoma. Oncotarget 6, 24 132–24 147. (doi:10.18632/oncotarget.4432)

56. Lu G, Shan T, He S, Ren M, Zhu M, Hu Y, Lu X, Zhang D. 2013 Overexpression of CENP-H as a novel prognostic biomarker for human

57. Su R, Nan H, Guo H, Ruan Z, Jiang L, Song Y, Nan K. 2016 Associations of components of PTEN/AKT/mTOR pathway with cancer stem cell markers and prognostic value of these biomarkers in hepatocellular carcinoma. Hepatol. Res. 46, 1380–1391. (doi:10.1111/hepr.12687)

58. Liu T, Chen J, Xiao S, Lei X. 2016 H2.0-like homeobox 1 acts as a tumor suppressor in hepatocellular carcinoma. Tumour Biol. 37, 6419–6428. (doi:10.1007/s13277-015-4490-z)

59. Huang DW, Sherman BT, Lempicki RA. 2009 Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57. (doi:10.1038/nprot.2008.211)

60. Mi H et al. 2017 PANTHER version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45(Database issue), D183–D189. (doi:10.1093/nar/gkw1138)

61. Stumpff J, Von Dassow G, Wagenbach M, Asbury C, Wordeman L. 2008 The kinesin-8 motor, Kif18A, suppresses kinetochore movements to control mitotic chromosome alignment. Dev. Cell 14, 252–262. (doi:10.1016/j.devcel.2007.11.014)

62. Kasahara M, Nagahara M, Nakagawa T, Ishikawa T, Sato T, Uetake H. 2016 Clinicopathological relevance of kinesin family member 18A expression in invasive breast cancer. Oncol. Lett. 12, 1909–1914. (doi:10.3892/ol.2016.4823)

63. Manchado E, Guillamot M, Malumbres M. 2012 Killing cells by targeting mitosis. Cell Death Differ. 19, 369–377. (doi:10.1038/cdd.2011.197)

64. Inoue M, Shinohara ML. 2011 Intracellular osteopontin (iOPN) and immunity. Immunol. Res. 49, 160–172. (doi:10.1007/s12026-010-8179-5)

65. Wang KX, Denhardt DT. 2008 Osteopontin: role in immune regulation and stress responses. Cytokine Growth Factor Rev. 19, 333–345. (doi:10.1016/j.cytogfr.2008.08.001)

66. Wen Y, Jeong S, Xia Q, Kong X. 2016 Role of osteopontin in liver diseases. Int. J. Biol. Sci. 12, 1121–1128. (doi:10.7150/ijbs.16445)

67. Qin L. 2014 Osteopontin is a promoter for hepatocellular carcinoma metastasis: a summary of 10 years of studies. Front. Med. 8, 24–32. (doi:10.1007/s11684-014-0312-8)

68. Wan HG, Xu H, Gu YM, Wang H, Xu W, Zu MH. 2014 Comparison osteopontin vs AFP for the diagnosis of HCC: a meta-analysis. Clin. Res. Hepatol. Gastroenterol. 38, 706–714. (doi:10.1016/j.clinre.2014.06.008)

69. Fouad SA, Mohamed NA, Fawzy MW, Moustafa DA. 2015 Plasma osteopontin level in chronic liver disease and hepatocellular carcinoma. Hepat. Mon. 15, e30753. (doi:10.5812/hepatmon.30753)

70. Weber GF. 2011 The cancer biomarker osteopontin: combination with other markers. Cancer Genomics Proteomics 8, 263–288.

71. Cao DX et al. 2012 Osteopontin as potential biomarker and therapeutic target in gastric and

liver cancers. *World J. Gastroenterol.* **18**, 3923–3930. (doi:10.3748/wjg.v18.i30.3923)

72. Boyault S *et al.* 2007 Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* **45**, 42–52. (doi:10.1002/hep.21467)

73. Villanueva A *et al.* 2011 Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma. *Gastroenterology* **140**, 1501–1512. (doi:10.1053/j.gastro.2011.02.006)

74. Li F, Ambrosini G, Chu EY, Plescia J, Tognin S, Marchisio PC, Altieri DC. 1998 Control of apoptosis and mitotic spindle checkpoint by survivin. *Nature* **396**, 580–584. (doi:10.1038/25141)

75. Mita AC, Mita MM, Nawrocki ST, Giles FJ. 2008 Survivin: key regulator of mitosis and apoptosis and novel target for cancer therapeutics. *Clin. Cancer Res.* **14**, 5000–5005. (doi:10.1158/1078-0432.CCR-08-0746)

76. Fortugno P, Beltrami E, Plescia J, Fontana J, Pradhan D, Marchisio PC, Sessa WC, Altieri DC. 2003 Regulation of survivin function by Hsp90. *Proc. Natl Acad. Sci. USA* **100**, 13 791–13 796. (doi:10.1073/pnas.2434345100)

77. Zhang LQ, Wang J, Jiang F, Xu L, Liu FY, Yin R. 2012 Prognostic value of survivin in patients with non-small cell lung carcinoma: a systematic review with meta-analysis. *PLoS ONE* **7**, e34100. (doi:10.1371/journal.pone.0034100)

78. Khan Z, Khan AA, Yadav H, Prasad GBKS, Bisen PS. 2017 Survivin, a molecular target for therapeutic interventions in squamous cell carcinoma. *Cell. Mol. Biol. Lett.* **22**, 8. (doi:10.1186/s11658-017-0038-0)

79. Nigam J, Chandra A, Kazmi HR, Parmar D, Singh D, Gupta V. 2014 Expression of survivin mRNA in gallbladder cancer: a diagnostic and prognostic marker? *Tumour Biol.* **35**, 9241–9246. (doi:10.1007/s13277-014-2200-x)

80. Liu JL, Zhang XJ, Zhang Z, Zhang AH, Wang W, Dong JH. 2013 Meta-analysis: prognostic value of survivin in patients with hepatocellular carcinoma. *PLoS ONE* **8**, e83350. (doi:10.1371/journal.pone.0083350)

81. Altieri DC. 2013 Targeting survivin in cancer. *Cancer Lett.* **332**, 225–228. (doi:10.1016/j.canlet.2012.03.005)

82. Holmes D. 2012 Cancer drug's survivin suppression called into question. *Nat. Med.* **18**, 842–843. (doi:10.1038/nm0612-842b)

83. Li D *et al.* 2016 Hepatocellular carcinoma repression by TNFα-mediated synergistic lethal effect of mitosis defect-induced senescence and cell death sensitization. *Hepatology* **64**, 1105–1120. (doi:10.1002/hep.28637)

84. Cleveland DW, Mao Y, Sullivan KF. 2003 Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell* **112**, 407–421. (doi:10.1016/S0092-8674(03)00115-6)

85. Zhang W, Mao JH, Zhu W, Jain AK, Liu K, Brown JB, Karpen GH. 2016 Centromere and kinetochore gene misexpression predicts cancer patient survival and response to radiotherapy and chemotherapy. *Nat. Commun.* **7**, 12 619. (doi:10.1038/ncomms12619)

86. Mishra PK, Au WC, Choy JS, Kuich PH, Baker RE, Foltz DR, Basrai MA, Sullivan BA. 2011 Misregulation of Scm3p/HJURP causes chromosome instability in *Saccharomyces cerevisiae* and human cells. *PLoS Genet.* **7**, e1002303. (doi:10.1371/journal.pgen.1002303)

87. Bakhoum SF, Compton DA. 2012 Kinetochores and disease: keeping microtubule dynamics in check! Current opinion in cell biology. **24**, 64–70. (doi:10.1016/j.ceb.2011.11.012)

88. Liao WT, Wang X, Xu LH, Kong QL, Yu CP, Li MZ, Shi L, Zeng M-S, Song L-B. 2009 Centromere protein H is a novel prognostic marker for human nonsmall cell lung cancer progression and overall patient survival. *Cancer* **115**, 1507–1517. (doi:10.1002/cncr.24128)

89. Tomonaga T, Matsushita K, Ishibashi M, Nezu M, Shimada H, Ochiai T, Yoda K, Nomura F. 2005 Centromere protein H is up-regulated in primary human colorectal cancer and its overexpression induces aneuploidy. *Cancer Res.* **65**, 4683–4689. (doi:10.1158/0008-5472.CAN-04-3613)

90. Liao WT *et al.* 2007 Centromere protein H is a novel prognostic marker for nasopharyngeal carcinoma progression and overall patient survival. *Clin. Cancer Res.* **13**, 508–514.

91. Zucman-Rossi J, Villanueva A, Nault JC, Llovet JM. 2015 Genetic landscape and biomarkers of hepatocellular carcinoma. *Gastroenterology* **149**, 1226–1239. (doi:10.1053/j.gastro.2015.05.061)

92. Bollard J *et al.* 2017 Palbociclib (PD-0332991), a selective CDK4/6 inhibitor, restricts tumour growth in preclinical models of hepatocellular carcinoma. *Gut* **66**, 1286. (doi:10.1136/gutjnl-2016-312268)

93. Lv J, Liu H, Wang Q, Tang Z, Hou L, Zhang B. 2003 Molecular cloning of a novel human gene encoding histone acetyltransferase-like protein involved in transcriptional activation of hTERT. *Biochem. Biophys. Res. Commun.* **311**, 506–513. (doi:10.1016/j.bbrc.2003.09.235)

94. Zhang H *et al.* 2014 GSK-3beta-regulated N-acetyltransferase 10 is involved in colorectal cancer invasion. *Clin. Cancer Res.* **20**, 4717–4729. (doi:10.1158/1078-0432.CCR-13-3477)

95. Ma R, Chen J, Jiang S, Lin S, Zhang X, Liang X. 2016 Up regulation of NAT10 promotes metastasis of hepatocellular carcinoma cells through epithelial-to-mesenchymal transition. *Am. J. Transl. Res.* **8**, 4215–4223.

96. Liu X *et al.* 2016 NAT10 regulates p53 activation through acetylating p53 at K120 and ubiquitinating Mdm2. *EMBO Rep.* **17**, 349–366. (doi:10.15252/embr.201540505)

97. Larrieu D, Britton S, Demir M, Rodriguez R, Jackson SP. 2014 Chemical inhibition of NAT10 corrects defects of laminopathic cells. *Science* **344**, 527–532. (doi:10.1126/science.1252651)

98. Andrisani OM, Studach L, Merle P. 2011 Gene signatures in hepatocellular carcinoma (HCC). *Semin. Cancer Biol.* **21**, 4–9. (doi:10.1016/j.semcancer.2010.09.002)

99. Zhao R, Li L, Yang J. 2016 PKM2 affects the development of hepatocellular carcinoma. *Int. J. Clin. Exp. Med.* **9**, 9744–9751.

100. Kent LN *et al.* 2017 Dosage-dependent copy number gains in E2f1 and E2f3 drive hepatocellular carcinoma. *J. Clin. Invest.* **127**, 830–842. (doi:10.1172/JCI87583)

101. Tanaka S, Arii S. 2012 Molecular targeted therapies in hepatocellular carcinoma. *Semin. Oncol.* **39**, 486–492. (doi:10.1053/j.seminoncol.2012.05.005)

102. Fernando J, Sancho P, Fernández-Rodriguez CM, Lledó JL, Caja L, Campbell JS, Fausto N, Fabregat I. 2012 Sorafenib sensitizes hepatocellular carcinoma cells to physiological apoptotic stimuli. *J. Cell. Physiol.* **227**, 1319–1325. (doi:10.1002/jcp.22843)

103. Cervello M, Bachvarov D, Lampiasi N, Cusimano A, Azzolina A, McCubrey JA, Montalto G. 2012 Molecular mechanisms of sorafenib action in liver cancer cells. *Cell Cycle* **11**, 2843–2855. (doi:10.4161/cc.21193)

104. Nagy A, Lanczky A, Menyhart O, Gyorffy B. 2018 Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci. Rep.* **8**, 9227. (doi:10.1038/s41598-018-27521-y)