ISSN 1993-6842 (on-line); ISSN 0233-7657 (print)
Biopolymers and Cell. 2016. Vol. 32. N 6. P 418–432
doi: http://dx.doi.org/10.7124/bc.000939

UDC 57.032

Transcriptome-based identification of PDGFA as a candidate secreted biomarker for hepatocellular carcinoma

M. S. Chesnokov[1], O. M. Krivtsova[1,2], P. A. Skovorodnikova[1,2], A. S. Makarova[1,2], I. F. Kustova[1], M. D. Logacheva[2], A. A. Penin[2], A. V. Klepikova[2], D. A. Shavochkina[1], N. E. Kudashkin[1], E. A. Moroz[1], Y. I. Patyutko[1], E. A. Kotelnikova[3], N. L. Lazarevich[1,2]

[1] N. N. Blokhin Russian Cancer Research Centre, RAMS
24, Kashirskoye shosse, Moscow, Russian Federation, 115478

[2] M. V. Lomonosov Moscow State University
Leninskie Gory, 1/12, Moscow, Russian Federation, 119991

[3] ZAO Personal Biomedicine,
124/17, Prospekt Mira, Moscow, Russian Federation, 129164
lazarevich.nl@gmail.com

Aim. Identification of candidate secreted biomarkers for hepatocellular carcinoma (HCC) diagnosis. **Methods.** Genes upregulated in HCC tissue and encoding secreted proteins were identified by RNA-seq. Gene expression changes in HCC were evaluated by RT-qPCR and meta-analysis of public databases. Biomarker properties were studied using ROC-curves, correlation and survival analysis. **Results.** PDGFA was identified by RNA-seq as an overexpressed gene encoding for a secreted protein in 5 HCC cases. PDGFA and GPC3 upregulation was revealed in 17 of 19 HCC samples and in most cases from the public databases. Combination of PDGFA and GPC3 discerned HCC and non-tumor tissue better than PDGFA or GPC3 alone. PDGFA overexpression was associated with better overall survival of the patients at early HCC stage and with weaker tumor invasion into blood vessels. **Conclusion.** PDGFA is a valuable secreted biomarker for HCC that might be used in combination with GPC3 to increase its sensitivity.

Keywords: hepatocellular carcinoma, PDGFA, tumor biomarker, NGS

Introduction

Hepatocellular carcinoma (HCC) is the most common form of malignant liver tumors with extremely high aggressiveness and poor prognosis. HCC ranks the second place in cancer-related mortality rates while most HCC patients are diagnosed at advanced stages when the existing therapeutic approaches become inefficient [1,2].

The major difficulty in improving HCC diagnosis and treatment is imposed by a high heterogeneity of the genetic and signaling aberrations observed in HCC and poor understanding of molecular mechanisms underlying its development. Thus, the identification of new biomarkers suitable for an early diagnosis and potential therapeutic targets is an important field in improving the efficiency of HCC management [3].

Alpha-fetoprotein (AFP), the only HCC marker approved for clinical practice, has low sensitivity for early tumor detection [4,5]. Among additional HCC biomarkers under investigation, glypican-3 (GPC3) is the most promising one that demonstrates high sensitivity and specificity in tumor tissue but performs worse when detected in blood serum. The efficiency of HCC diagnosis can be improved by using

© 2016 M. S. Chesnokov *et al.*; Published by the Institute of Molecular Biology and Genetics, NAS of Ukraine on behalf of Biopolymers and Cell.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited

combinations of biomarkers but remains insufficient to confidently detect HCC at early stages [5,6].

The next-generation sequencing (NGS) approaches open new possibilities in disclosing the molecular basis of carcinogenesis. The genomic and transcriptomic data revealed the multiple tumor-specific mutations and gene expression changes that can be further analyzed to identify putative biomarkers and changes in the signaling pathways regulating HCC progression. The present work is devoted to identification of novel prospective HCC biomarkers based on the results of transcriptome sequencing and investigation of their potential impact using experimental and bioinformatic approaches.

Materials and Methods

Samples collection, RNA extraction, transcriptome sequencing and differential expression analysis

19 pairs of tumor and adjacent non-tumorous (NT) liver tissues were collected after tumor resection from the patients with histologically verified HCC not associated with hepatitis virus infection. The samples were collected with informed consent, conforming to the ethical guidelines of the 1975 Declaration of Helsinki, frozen in liquid nitrogen and stored at –80 °C. The clinicopathological data on collected cases are presented in Table 1.

Total RNA was isolated as previously described [7]. Illumina HiSeq2000 100 nt pair-end transcriptome sequencing was performed for 5 pairs of tumor and liver tissue in two biological replicates. Library preparation, transcriptome sequencing, read processing and differential expression analysis were performed as previously described [7]**.**

Table 1. **Clinicopathological data on HCC patients enrolled in present study.**

Characteristics	Number of cases (n=19)
Age, years (mean±SD)	48.1±19.2
Gender, male/female	11/8
TNM staging, I/II/III/IV	4/4/6/5
Tumor size, cm (mean ± SD)	10.1±5.6
Tumor capsule, absent/feeble/prominent/N/A[1]	4/9/5/1
Blood vessel invasion, yes/no	11/8
Tumor vascularity, low/moderate/high/N/A	2/5/5/7
Histological differentiation, Edmondson-Steiner grade, G1/G2/G3/Gx[2]	3/8/3/5
Intrahepatic metastases, yes/no	8/11
Lymph node metastases, yes/no	3/16
Distant metastases, yes/no	1/18
AFP serum level, low (<50 ng/ml)/high (>50 ng/ml)	9/9
Cirrhosis, yes/no	5/14
Tumor necrosis, yes/no	12/7

[1] N/A – data not available

[2] Gx – Edmondson-Steiner grade not applicable

Quantitative Real-Time PCR

Total RNA was reverse transcribed using random hexanucleotide primers and MMLV reverse transcriptase (Promega, USA).

Real-time RT-qPCR was carried out using SYBR Green I PCR kit (Syntol, Russian Federation) and iQ5 Multicolor Real-Time PCR Detection System (Bio-Rad Laboratories, USA). TATA-binding protein gene (*TBP*) was used as a reference gene. 45 cycles of amplification (30 s at 95 °C, 30 s at annealing temperature (*PDGFA* – 66.0 °C, *GPC3* – 67.7 °C, *TBP* – 62.8 °C), 30s at 72 °C) were performed and the reaction specificity was checked afterwards by a melt curve analysis. The gene expression levels were estimated using a standard curve for fixed signal value. For each sample, the gene expression level was normalized to TBP expression, logarithm to base 2 was taken from normalized value and difference between the values obtained for HCC and corresponding NT samples was calculated.

The following primers were used for reactions: *PDGFA*-forward 5'-ACCACCGCAGCGTCAAGG-3', *PDGFA*-reverse 5'-GCGGCTCATCCTCACCTC AC-3', *GPC3*-forward 5'-GCAGGAAAGCTGACC ACCAC-3', *GPC3*-reverse 5'-AGTTCCCTTCTTC GGCTGGAT-3', *TBP*-forward5'-TGCACAGGAGC CAAGAGTGA-3', *TBP*-reverse 5'-ACTTCACATC ACAGCTCCCCA-3'.

Table 2. **MIQE qPCR information table.**

Item to check	Importance	Information
EXPERIMENTAL DESIGN		
Definition of experimental and control groups	E	Experimental group consisted of HCC tissue samples and control group consisted of corresponding samples of non-tumorous liver tissue taken from same patients
Number within each group	E	19 samples within each group
Assay carried out by the core or investigator's laboratory?	D	Investigator's laboratory
Acknowledgment of authors' contributions	D	
SAMPLE		
Description	E	Tissue samples of HCC and non-tumorous liver tissue taken from patients diagnosed with HCC after tumor resection
Volume/mass of sample processed	D	Approximately 20 mg of tissue
Microdissection or macrodissection	E	Macrodissection
Processing procedure	E	Tissue samples were taken immediately after tumor resection with sterile scalpel and incubated in RNAlater Stabilization Solution (Thermo Fisher Scientific AM7024, USA) for 24 hours at 4 °C. Then RNAlater Stabilization Solution was removed and samples were frozen
If frozen, how and how quickly?	E	Samples were frozen at –70 °C immediately after RNAlater Stabilization Solution treatment
If fixed, with what and how quickly?	E	Samples were not fixed
Sample storage conditions and duration	E	Frozen samples were stored at –70 °C
NUCLEIC ACID EXTRACTION		
Procedure and/or instrumentation	E	Frozen samples were homogenized using glass Potter grinder cooled with liquid nitrogen and transferred into RNase-free tubes
Name of kit and details of any modifications	E	RNA was extracted from homogenized samples using PureLink RNA Mini Kit (Thermo Fisher Scientific 12183018A, USA) with additional on-column DNase treatment according to manufacturer's manual
Source of additional reagents used	D	Thermo Fisher Scientific, USA
Details of DNase or RNase treatment	E	On-column DNase treatment was performed with PureLink DNase Set (Thermo Fisher Scientific 12185010, USA) according to manufacturer's manual
Contamination assessment (DNA or RNA)	E	Contamination of RNA samples by protein or organic chemical compounds was evaluated by measuring optical density at 260, 280 and 230 nm. Contamination of RNA samples by genomic DNA was evaluated by using control sample without reverse transcriptase added during reverse transcription step
Nucleic acid quantification	E	RNA concentration was evaluated by measuring optical density at 260 nm
Instrument and method	E	NanoDrop ND-1000 spectrophotometer (Thermo Scientific, USA)
Purity (A_{260}/A_{280})	D	All examined RNA samples had A_{260}/A_{280} ratio higher than 2.0
Yield	D	Approximately 20 μg of total RNA were extracted from one tissue sample

Item to check	Importance	Information
RNA integrity: method/ instrument	E	RNA integrity was evaluated using Agilent 2100 Bioanalyzer (Agilent Technologies, USA)
RIN/RQI or Cq of 3' and 5' transcripts	E	Only samples with RIN (RNA integrity number) > 7 were taken for analysis
Electrophoresis traces	D	Electrophoresis was not used
Inhibition testing (Cq dilutions, spike, or other)	E	Standard calibration curve was used for evaluating possible PCR inhibitor contamination; PCR efficiency of 98 %–102 % was considered applicable
REVERSE TRANSCRIPTION		
Complete reaction conditions	E	Total RNA was reverse transcribed using random hexanucleotide primers and RevertAid Reverse Transcriptase. Mix of hexanucleotide primers and total RNA template (total volume 12.25 µl) was incubated at 70 °C for 10 min and cooled at 4 °C. After that, 7.75 µl of reaction mix containing RevertAid Reaction buffer for RT, dNTPs, DTT, Ribo-Lock RNase inhibitor and RevertAid Reverse Transcriptase was added. Final concentrations of reagents were 5 ng/µl hexanucleotide primers, 0.25 mM each dNTP, 100 µM DTT, 1 U/µl Ribo-Lock and 2.5 U/µl RevertAid RT. Reaction mix was incubated at 42 °C for 60 min, at 95 °C for 10 min and cooled at 4 °C.
Amount of RNA and reaction volume	E	2 µg of total RNA was used for RT reaction. Reaction volume for incubation of RNA with hexanucleotide primers was 12.25 µl and reaction volume for RT reaction was 20 µl
Priming oligonucleotide (if using GSP) and concentration	E	Random hexanucleotide primers, 5 ng/µl
Reverse transcriptase and concentration	E	RevertAid Reverse Transcriptase, 2.5 U/µl
Temperature and time	E	See "Complete reaction conditions" subsection
Manufacturer of reagents and catalogue numbers	D	Random hexanucleotide primers (Syntol, Russian Federation) dNTPs (Syntol dNTP-100-010, Russian Federation) DTT (Sigma-Aldrich D9779-10G, USA) Ribo-Lock RNase inhibitor (Thermo Fisher Scientific EO0381, USA) RevertAid Reverse Transcriptase (Thermo Fisher Scientific EP0441, USA) RevertAid 5X Reaction buffer for RT (Thermo Fisher Scientific EP0441, USA)
C_qs with and without reverse transcription	D	C_q for experimental samples was within the range of 20–35 cycles, no amplification or very low non-specific amplification was observed in samples without reverse transcription (C_q>40 cycles)
Storage conditions of cDNA	D	Resulting cDNA samples were diluted up to 100 µl by double-distilled water and stored at –20 °C
qPCR TARGET INFORMATION		
Gene symbol	E	*TBP*; *PDGFA*; *GPC3*
Sequence accession number	E	TBP - NM_003194.4 PDGFA - NM_002607.5 GPC3 - NM_001164617.1
Location of amplicon	D	Amplicons were located in conservative regions of target gene mRNAs to amplify all possible isoforms
Amplicon length	E	TBP – 136 b.p. PDGFA – 166 b.p. and 235 b.p. GPC3 – 161 b.p. and 230 b.p. (not detected)

Item to check	Importance	Information
In silico specificity screen (BLAST, and so on)	E	All primers were checked for specificity using NCBI Primer-BLAST service (https://www.ncbi.nlm.nih.gov/tools/primer-blast/)
Pseudogenes, retropseudogenes, or other homologs?	D	Primers were designed to prevent amplification of homologous templates
Sequence alignment	D	Not evaluated
Secondary structure analysis of amplicon	D	Not evaluated
Location of each primer by exon or intron (if applicable)	E	*TBP*-forward – exon 5 of NM_003194.4 transcript *TBP*-reverse – exon 6 of NM_003194.4 transcript *PDGFA*-forward – exon 4 of NM_002607.5 transcript *PDGFA*-reverse – exon 7 of NM_002607.5 transcript *GPC3*-forward – exon 3 5 of NM_004484.3 transcript *GPG3*-reverse – exon 5 of NM_004484.3 transcript
What splice variants are targeted?	E	*TBP* - NM_001172085, NM_003194 *PDGFA* - NM_002607, NM_033023 *GPC3* - NM_001164617, NM_001164618, NM_001164619, NM_004484
qPCR OLIGONUCLEOTIDES		
Primer sequences	E	*TBP*-forward – 5'-TGCACAGGAGCCAAGAGTGA-3' *TBP*-reverse – 5'-ACTTCACATCACAGCTCCCCA-3' *PDGFA*-forward – 5'-ACCACCGCAGCGTCAAGG-3' *PDGFA*-reverse – 5'-GCGGCTCATCCTCACCTCAC-3' *GPC3*-forward – 5'-GCAGGAAAGCTGACCACCAC-3' *GPC3*-reverse – 5'-AGTTCCCTTCTTCGGCTGGAT-3'
RTPrimerDB identification number	D	Not applicable
Probe sequences	D	Not applicable
Location and identity of any modifications	E	Not applicable
Manufacturer of oligonucleotides	D	Syntol, Russian Federation
Purification method	D	PAAG electrophoresis
qPCR PROTOCOL		
Complete reaction conditions	E	PCR was performed using SYBR Green I PCR kit (Syntol R-402, Russian Federation). Reaction mix contained cDNA template, PCR buffer with SYBR Green I dye, dNTPs, $MgCl_2$, oligonucleotide primers and SynTaq DNA polymerase. Following PCR protocol was used: Initial denaturation – 5 min at 95 °C 45 PCR cycles – 30 sec at 95 °C, 30 sec at primer annealing temperature (PDGFA – 66.0 °C, GPC3 – 67.7 °C, TBP – 62.8 °C), 30 sec at 72 °C Final elongation – 3 min at 72 °C Melt curve analysis – from 60 °C up to 100 °C with increment of 0.5 °C
Reaction volume and amount of cDNA/DNA	E	PCR was performed in a volume of 25 µl, cDNA sample (see "Storage conditions of cDNA" subsection) was diluted ten-times and 10 µl of diluted cDNA sample were added to the reaction mix
Primer, (probe), Mg^{2+}, and dNTP concentrations	E	0.4 µM of each primer, 2.5 µM Mg^{2+}, 0.25 mM dNTPs

Item to check	Importance	Information
Polymerase identity and concentration	E	SynTaq DNA polymerase (Syntol E-039-1000, Russian Federation)
Buffer/kit identity and manufacturer	E	SYBR Green I PCR kit (Syntol R-402, Russian Federation).
Exact chemical composition of the buffer	D	Not specified by manufacturer
Additives (SYBR Green I, DMSO, and so forth)	E	SYBR Green I dye was pre-added to PCR buffer by manufacturer
Manufacturer of plates/tubes and catalog number	D	0.2 ml PCR strips with domed caps (SSI 3240-00, USA)
Complete thermocycling parameters	E	See "Complete reaction conditions" subsection
Reaction setup (manual/robotic)	D	Manual
Manufacturer of qPCR instrument	E	iCycler Thermal Cycler with iQ5 Multicolor Real-Time PCR Detection System, data were analyzed using iQ5 Optical System Software (Bio-Rad Laboratories, USA)
qPCR VALIDATION		
Evidence of optimization (from gradients)	D	Primer annealing temperature was optimized for each primer set using qPCR with temperature gradient to achieve reaction efficiency of 98 %-102 % and minimize non-specific amplification
Specificity (gel, sequence, melt, or digest)	E	Specificity of qPCR was determined using melt curve analysis
For SYBR Green I, C_q of the NTC	E	No amplification or very low non-specific amplification was observed in samples without reverse transcription and in NTC samples (C_q>40 cycles)
Calibration curves with slope and y intercept	E	Slope and y intercept values: *TBP* – slope=-3.283, y int=31.467 *PDGFA* – slope=-3.339, y int=30.015 *GPC3* – slope=-3.301, y int=26.564
PCR efficiency calculated from slope	E	*TBP* – E=101.6 % *PDGFA* – E=99.3 % *GPC3* – E=100.9 %
CIs for PCR efficiency or SE	D	Not evaluated
r^2 of calibration curve	E	*TBP* – E=0.996 *PDGFA* – E=0.998 *GPC3* – E=0.998
Linear dynamic range	E	Standard curve was linear within the limits of cDNA dilutions from non-diluted samples up to 1/10000 dilution for all primers used
C_q variation at LOD	E	SD of C_q values for 1/10000 dilution sample: *TBP* – 0.83 *PDGFA* – 0.77 *GPC3* – 0.37
CIs throughout range	D	Not evaluated
Evidence for LOD	E	High variation of C_q values obtained for 1/10000 dilution sample
If multiplex, efficiency and LOD of each assay	E	Displayed above

Item to check	Importance	Information
DATA ANALYSIS		
qPCR analysis program (source, version)	E	Data were analyzed using iQ5 Optical System Software, v2.0 (Bio-Rad Laboratories, USA)
Method of C_q determination	E	C_q values were determined by setting a signal threshold. Threshold value was adjusted to detect the amplification of samples during log-phase of PCR and at the same time to be significantly higher than background signal
Outlier identification and disposition	E	Data were checked for outliers using Grubbs' test, no significant outliers were detected
Results for NTCs	E	No amplification or very low non-specific amplification was observed in samples without reverse transcription and in NTC samples (C_q>40 cycles)
Justification of number and choice of reference genes	E	*TBP* was chosen as a reference gene based on lowest variability of its expression level in 10 samples (5 HCC samples and 5 corresponding non-tumorous tissue samples) examined by RNAseq approach. Low *TBP* expression variability was validated on TCGA-LIHC RNAseq dataset (see Materials and Methods). This experimental evidence is supported by previously published data [9, 10] while the expression of other reference gene, HPRT was found to be variable in some non-HBV related HCC samples.
Description of normalization method	E	For each sample, target gene expression was normalized to *TBP* expression. Level of target gene expression was divided by level of *TBP* expression prior to comparing different samples
Number and concordance of biological replicates	D	Not applicable since one HCC tissue and one non-tumorous tissue samples were taken from each patient
Number and stage (reverse transcription or qPCR) of technical replicates	E	At least 3 technical replicates were analyzed in each qPCR reaction
Repeatability (intraassay variation)	E	Average CV for technical replicates was 0.22
Reproducibility (interassay variation, CV)	D	Not evaluated
Power analysis	D	Not evaluated
Statistical methods for results significance	E	Changes in expression level that were higher than 2-fold were considered significant when comparing HCC and non-tumorous tissue samples from same patient. Differences between gene expression levels in HCC and NT sample sets were estimated using paired sample sign test (for sets of paired samples) and Mann-Whitney U-test (for sets of unpaired samples).
Software (source, version)	E	Origin Pro 2016 software (OriginLab Corporation, USA)
C_q or raw data submission with RDML	D	Not submitted

All relevant qPCR conditions and characteristics determined by the MIQE Guidelines [8] are described in Table 2.

Hepatocellular carcinoma datasets

Publicly available datasets containing information on gene expression in paired liver-HCC samples and their clinical data were acquired from GEO database (https://www.ncbi.nlm.nih.gov/geo): GSE14520 [11], GSE25599 [12], GSE5364 [13], GSE65485 [14], GSE77314 [15]. PDGFA expression data generated using 205463_s_at probe set were considered for GSE14520 as it targets both PDGFA isoforms.

TCGA Liver Hepatocellular Carcinoma (TCGA-LIHC) set that comprised information on the normalized gene expression in 51 matched liver and tumor

tissues and 321 tumors corresponding to a pooled normal sample was obtained from The Cancer Genome Atlas Network (https://gdc-portal.nci.nih.gov/) through FireBrowse (http://www.firebrowse.org). The clinical and survival data were downloaded at cBioPortal (http://www.cbioportal.org).

Statistical analysis

Each tissue sample used for RT-qPCR was analyzed in at least four technical replicates and a mean value was used for further analysis. The statistical analysis of results and plotting of graphs were performed using Origin Pro 2016 software (OriginLab Corporation, USA). The differences between gene expression levels in HCC and NT samples were estimated using a paired sample sign test (for paired samples) and Mann-Whitney U-test (for unpaired samples). The empirical distribution curves for gene expression in different datasets were compared using Kolmogorov-Smirnov test. The hierarchical cluster analysis of gene expression datasets was performed using Euclidean distance and Complete linkage algorithm. Receiver operating characteristic (ROC) curve discriminative power analysis was performed using the normalized gene expression level for classifier, HCC samples for estimation of true positive rate and NT samples from the same patients for estimation of false positive rate. The combined *PDGFA+GPC3* classifier for ROC curves was generated by applying a logistic regression model to the data on both *PDGFA* and *GPC3* expression levels and taking the values of expected probabilities as a new classifier. The correlations were evaluated using Spearman's rank correlation test. A survival analysis was performed using Kaplan-Meyer test with log-rank significance estimation algorithm. Statistical significance was accepted with $p<0.05$.

Results

The whole transcriptome data analysis using DESeq [16] revealed 83 differentially expressed (DE)

Table 3. **Potential secreted HCC markers that were significantly upregulated in 5 sequenced HCC cases**

Symbol (Entrez Gene ID)	Average fold change of gene expression	SignalP	TMHMM	Gene Cards	Protein Atlas	Literature	Average FPKM	
							Liver tissue	Tumor
VCAN (1462)	113.67 (14.61–261.68)	+	–	+	+	+	0.25 (0.20–0.30)	36.07 (7.39–74.95)
CCDC80 (151887)	27.02 (7.30–54.85)	+	–	+	+	+	0.32 (0.15–0.41)	9.63 (3.88–12.10)
COL1A1 (1277)	33.41 (19.85–49.81)	+	–	+	+	+	0.63 (0.37–0.93)	37.27 (13.55–52.08)
SMOC2 (64094)	20.32 (7.67–34.26)	+	–	+	+	+	0.15 (0.05–0.43)	3.07 (2.00–4.50)
LTBP2 (4053)	28.97 (11.58–56.28)	+	–	+	+	+	0.19 (0.10–0.29)	6.53 (2.89–10.52)
COL5A1 (1289)	9.92 (6.30–14.57)	+	–	+	+	+	0.50 (0.40–0.69)	8.16 (3.85–16.13)
PDGFA (5154)	18.99 (8.04–38.57)	+	–	+	+	+	0.14 (0.09–0.24)	4.19 (2.16–7.75)
ITGBL1 (9358)	16.38 (5.61–23.86)	+	–	+	+	+	0.30 (0.14–0.68)	5.48 (2.97–7.65)
COL15A1 (1306)	134.11 (30.05–314.65)	+	–	+	+	+	0.06 (0.04–0.09)	6.76 (3.00–14.23)

genes that were up-regulated more than 5-fold in all HCC samples as compared to corresponding adjacent liver tissue. In order to identify putative secreted HCC markers, the FASTA sequences of all mRNA isoforms of DE genes were analyzed with SignalP Server 4.1 (http://www.cbs.dtu.dk/services/SignalP/) and TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) using default settings. The genes harboring sequences that were predicted to encode the signal peptide cleavage sites but not transmembrane helices were examined using GeneCards (http://www.genecards.org/), The Human Protein Atlas (http://www.proteinatlas.org/) and the information from journal articles. Then we excluded the genes that were valuably expressed in normal liver (FPKM>1) and/or demonstrated relatively low level of expression in tumors (FPKM<2). Thus, we obtained a list comprising 9 potential secreted HCC markers (Table 3).

The list of candidate serum markers includes growth factor PDGFA, a component of PDGF signaling pathway identified in our recently published HCC

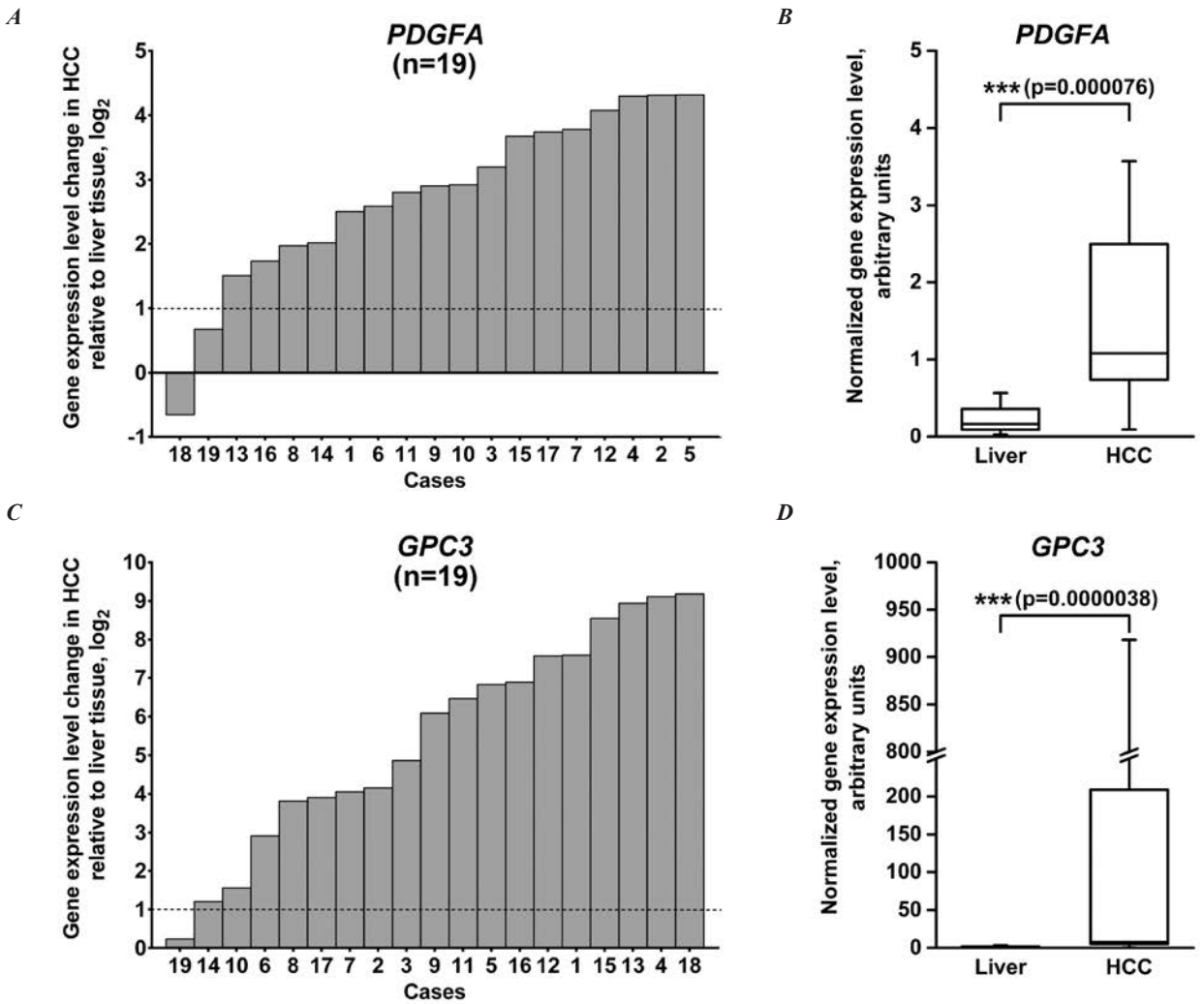

Fig. 1. Alteration of *PDGFA* (*A*, *B*) and *GPC3* (*C*, *D*) expression in examined HCC tissue samples. *A,* C — RT-qPCR analysis of *PDGFA* or *GPC3* expression changes in HCC specimens relative to matching non-tumor liver tissue for individual cases. *B, D* — Box plot representation of *TBP*-normalized expression levels of *PDGFA* and *GPC3* in examined sets of HCC and NT samples (n=19).

case report as a potential druggable target [7]. Since the proangiogenic and mitogenic stimulation promoted by PDGF signaling can be blocked by a multikinase inhibitor sorafenib, the only FDA approved drug for HCC treatment [17, 18], we have focused on investigation of the expression alterations of *PDGFA* that might be not only a candidate HCC marker but also a prospective target for drug treatment.

To explore HCC-specific changes in *PDGFA* expression we performed RT-qPCR analysis of *PDGFA* expression levels in 19 pairs of tumor and NT tissues from hepatitis-negative HCC patients. While low *PDGFA* expression levels were detected in all NT specimens, the *PDGFA* expression in HCC tissue was up-regulated more than two-fold in 17 of 19 (89.5 %) examined cases (Fig. 1A) and the difference between these two subsets was statistically significant (Fig. 1B).

We investigated the potential of *PDGFA* usage as a biomarker by comparing its expression changes in HCC tissue to the expression changes of *GPC3*, the latter being a promising candidate biomarker for HCC [19]. RT-qPCR analysis revealed the significant *GPC3* overexpression in HCC compared to NT tissue (Fig. 1D) in 18 of 19 (94.7 %) cases (Fig. 1C). Spearman's correlation analysis demonstrated that the changes in *PDGFA* expression were not associated with the clinicopathological properties of examined tumors.

In order to determine whether the *PDGFA* up-regulation discovered in the examined sample set is a frequent event in HCC we performed meta-analysis of the gene expression data for paired HCC/NT samples obtained from six publicly available datasets (Table 4). Each of the analyzed datasets displayed a significant (more than 2-fold) up-regulation of the *PDGFA* transcription in tumor tissue compared to the corresponding surrounding liver samples in no less than 50 % of cases. Since several datasets comprised the data on a low number of samples, we further analyzed TCGA ("TCGA set") and GSE14520 ("Roessler set") datasets.

The ratios of cases with significant *PDGFA* up-regulation ranged from 52.4 % ("Roessler set") to 63.8 % ("TCGA set") (Fig. 2A). Median values of normalized *PDGFA* expression level were significantly higher in cohorts of HCC tissue samples than in cohorts of corresponding NT specimens (Fig. 2B).

While both datasets support the observation of *PDGFA* up-regulation being a frequent event in HCC tissue, the percentage of *PDGFA* overexpressing samples is less than observed in our experimental set. To explore whether this difference could be associated with hepatitis infection we subtracted a fraction of 94 TCGA cases that were not marked as hepatitis-positive (hereinafter called "TCGA-HN set" for "hepatitis-negative"). The proportion of cases with up-regulated *PDGFA* expression in "TCGA-HN set" (68.1 %) (Fig. 2A) was very similar to the one observed in full "TCGA set", while no statistically significant differences between full and "HN" sets in the context of *PDGFA* expression level median values (p=0.558 estimated by Mann-Whitney U-test) and the empirical distribution curve (p=0.988 estimated by Kolmogorov-Smirnov test) were found thus indicating that *PDGFA* up-regulation in HCC occurs irrespectively of tumor etiology.

To evaluate the *PDGFA* potential sensitivity as a HCC biomarker we compared the *PDGFA* expression changes in "Roessler set" and "TCGA-HN set" to the alterations of *GPC3* expression (Fig. 3A). While the

Table 4. **Overexpression of *PDGFA* in HCC in six datasets comprising paired normal-tumor samples.**

Dataset ID	Number of cases	HBV-positive cases, %	*PDGFA* overexpression in tumor, % of cases	Gene expression profiling method
TCGA-LIHC	373	75	63.8	RNAseq
GSE14520	231	98	52.4	Affymetrix Human Genome HT U133A Array
GSE25599	10	100	50	RNAseq
GSE5364	8	N/A	100	Affymetrix HG U133A microarray
GSE77314	50	N/A	60	RNAseq
GSE65485	8	100	75	RNAseq

[1] N/A – data not available

sensitivity of *PDGFA* (52.4 % for "Roessler set", 68.1 % for "TCGA-HN set") was lower than that of *GPC3* (87.4 % for "Roessler set", 81.9 % for "TCGA-HN set"), the combination of *PDGFA* and *GPC3* increased the sensitivity to 93.9 % for "Roessler set" and 93.6 % for "TCGA-HN set" (p=0.024 for both sets compared to *GPC3* alone, estimated by Fisher's exact test).

In order to evaluate the possibility of using the expression level of *PDGFA*, *GPC3* or both genes as a parameter discerning HCC from NT tissue, we

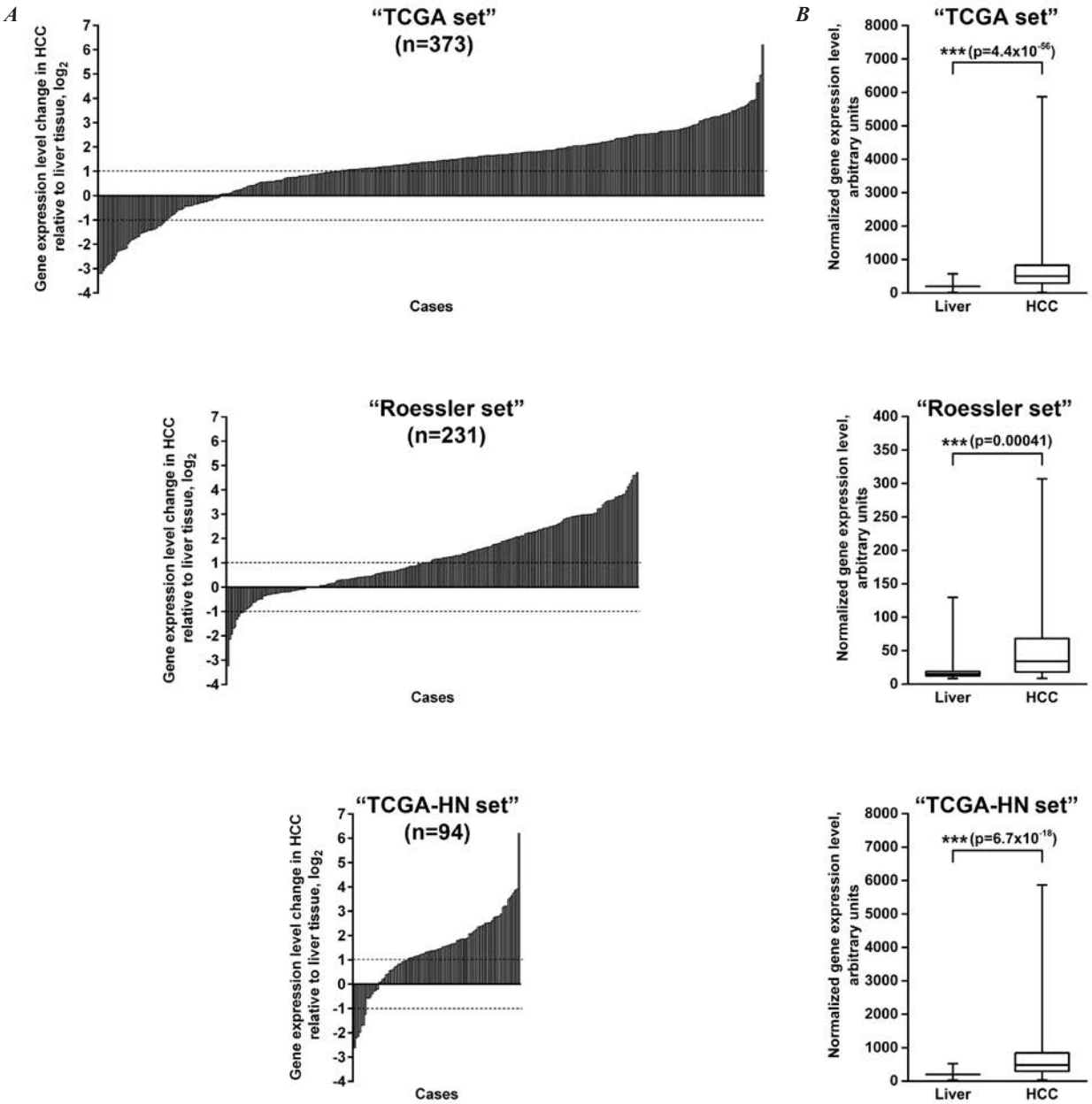

Fig. 2. Alteration of *PDGFA* expression observed in publicly available HCC datasets. *A* — Distribution curves of *PDGFA* expression level changes in individual HCC samples compared to NT samples. *B* — Box plot representation of normalized expression levels of *PDGFA* in examined sets of HCC and NT samples.

generated ROC curves using the data for paired samples from "Roessler set" (n=231) and "TCGA-HN set" (n=24). Usage of the *PDGFA+GPC3* combination increased the value of area under a curve (AUC) in comparison to *PDGFA* or *GPC3* alone, thus indicating a stronger discriminative power of the *PDGFA* and *GPC3* combination (Fig. 3B).

A correlation analysis of the *PDGFA* expression changes and clinicopathological characteristics available for "TCGA-HN set" revealed a reverse correlation of the *PDGFA* up-regulation with the extent of tumor invasion into blood vessels.

Since "Roessler set" contains the data on Barcelona Clinic Liver Cancer (BCLC) staging which is widely used for evaluation of prognosis and treatment algorithm for HCC patients [20] we analyzed an association between the *PDGFA* up-regulation and overall and progression-free survival of patients belonging to different BCLC groups. The *PDGFA* overexpression in tumor tissue was associated with better overall survival of patients with early BCLC-0 and BCLC-A HCC stages but not intermediate BCLC-B or late BCLC-C stage (Fig. 4). No associations between the *PDGFA* overexpression

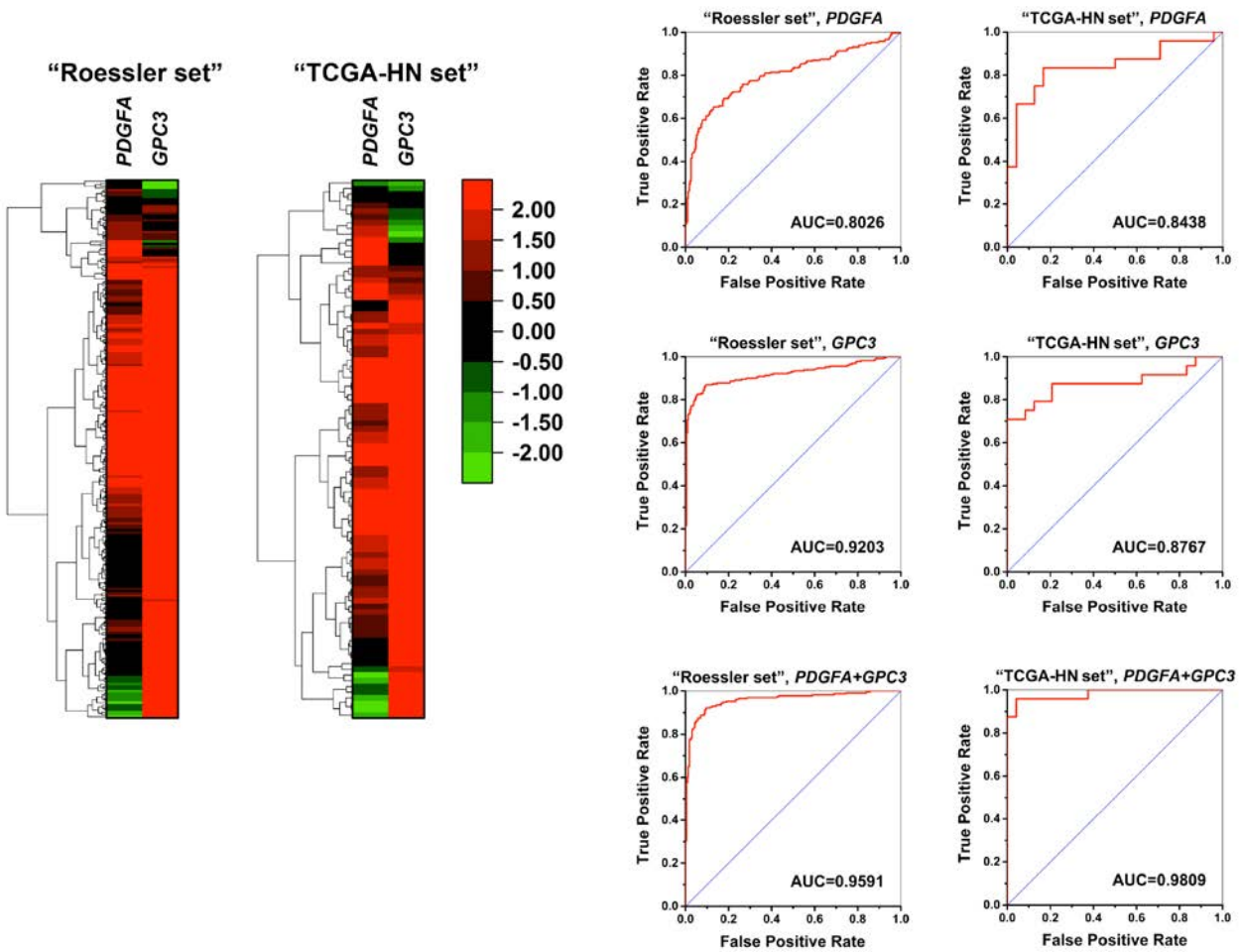

Fig. 3. Comparison of *PDGFA* and *GPC3* expression changes in HCC and NT tissue samples from publicly available datasets. *A* — Heatmap representation of *PDGFA* and *GPC3* expression changes in HCC specimens from "Roessler set" and "TCGA-HN set". Color bar indicates the alteration of gene expression in tumor tissue relative to NT sample in $\log_2$ scale. *B* — ROC curves representing the ability of *PDGFA*, *GPC3* or *PDGFA+GPC3* combination to discriminate HCC and NT tissue samples.

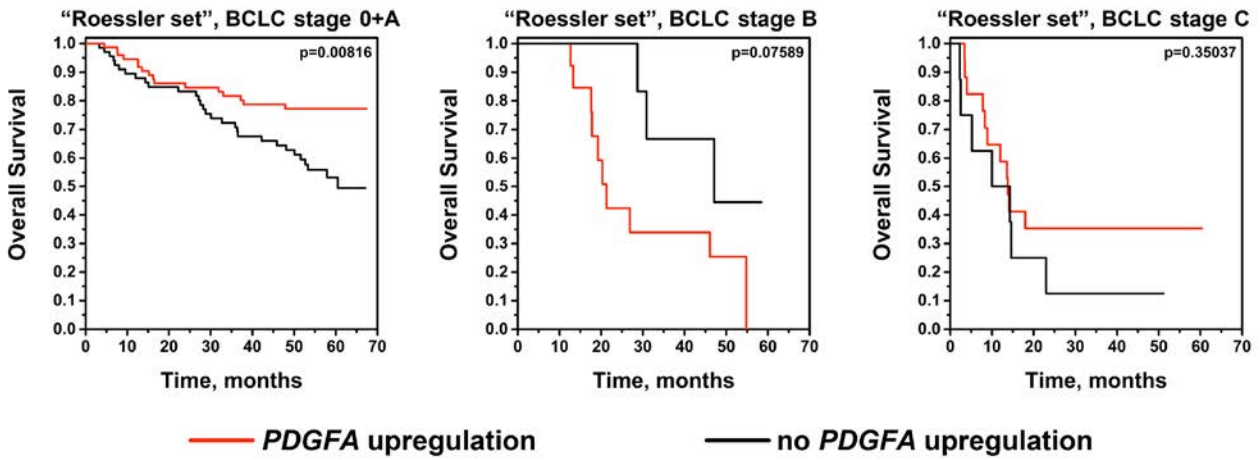

Fig. 4. Kaplan-Meyer analysis of association between *PDGFA* up-regulation and overall survival of patients from "Roessler set" belonging to different BCLC groups.

and progression-free survival were found (data not shown).

Discussion

The discovery of tumor biomarkers significantly improved the outcome for cancer patients and opened new possibilities for early diagnosis and targeted treatment of malignant tumors [21]. The only serum HCC biomarker approved for clinical practice is AFP [4] that displays 59% sensitivity and 90% specificity [22]. Since AFP exhibits insufficient sensitivity for the confident HCC diagnosis, the additional markers to complement AFP and improve HCC diagnostic accuracy are under investigation [23].

Currently GPC3 is considered to be one of the most promising HCC candidate biomarkers. It can be detected at the mRNA level in liver tissue or at the protein level in serum or liver tissue. Immunohistochemical detection of GPC3 demonstrates a high sensitivity for poorly-differentiated HCC but a lower sensitivity for highly-differentiated and fibrolamellar variants [19]. The *GPC3* mRNA was found to be overexpressed in more than 80 % of HCC cases associated with viral hepatitis and in 76 % of non-viral HCC cases [24]. However, the measuring of serum GPC3 level was less sensitive (55.2 %) while retaining a high specificity (84.2 %). The GPC3 combination with AFP was uncovered to be more effective for HCC diagnosis with 75.7 % sensitivity and 83.3 % specificity [25]. Thus, we have chosen GPC3 as a "reference" HCC biomarker and compared the data obtained for *PDGFA* to the *GPC3* performance.

Performed analysis of the expression data from our HCC set and publicly available databases revealed the frequent *PDGFA* overexpression in HCC tissue. Though *PDGFA* was previously reported to be overexpressed in HCC [26], no detailed investigation on its expression alteration or its potential as a HCC biomarker has been published to date.

A high rate of *PDGFA* up-regulation in 19 examined hepatitis-negative HCC cases was comparable to that of *GPC3*. However, *PDGFA* did not perform so well in larger and less homogenous datasets exhibiting a lower sensitivity than *GPC3*. While most cases from publicly available datasets demonstrated up-regulation of both *PDGFA* and *GPC3*, there were subsets with mutually exclusive overexpression of *PDGFA* or *GPC3* indicating that their combination could perform better than each biomarker separately. Indeed, if 2-fold increase in the expression level of either *PDGFA* or *GPC3* was taken as cut-off, the sensitivity of HCC detection considerably increased up to 93.6%. The analysis of biomarkers discriminatory power revealed that *PDGFA* and *GPC3*, when combined, distinguished HCC from NT tissue of the same patients better than *PDGFA* or *GPC3* individually.

PDGFA, a secretable protein detectable in patient's serum, may be considered as a potential HCC diagnostic marker at the mRNA or protein levels especially when used in combination with GPC3 to significantly improve its low sensitivity. The association of *PDGFA* up-regulation with better overall survival of the patients with BCLC-0 and BCLC-A early HCC stages and a weaker invasion of tumor cells into blood vessels demonstrates that it can be accounted as a prognostic factor. However, this putative prognostic impact is limited since it is not observed in the groups with BCLC stages B and C. Hence, the *PDGFA* up-regulation may be considered as a factor of favorable prognosis but the validation of this hypothesis requires further studies of larger patient cohorts.

Conclusion

The present study demonstrates that *PDGFA* is frequently overexpressed in HCC tissue. The combination of *PDGFA* and *GPC3* performs well in distinguishing HCC and NT tissue when detected at the mRNA level. *PDGFA* up-regulation might have a prognostic potential for the patients with early HCC stages. We suggest that PDGFA may be a promising HCC diagnostic biomarker. Further studies focused on the detection of PDGFA in tumor tissue and serum of the HCC patients are necessary to define its efficiency (either alone or in combination with other biomarkers) and the validity for improving sensitivity of the early HCC stages detection.

Funding

The work was partly supported by grant from Russian Ministry of Education and Science (contract 14.607.21.0049, RFMEFI60714X0049).

REFERENCES

1. *Colagrande S, Inghilesi AL, Aburas S, Taliani GG, Nardi C, Marra F.* Challenges of advanced hepatocellular carcinoma. *World J Gastroenterol.* 2016;**22**(34):7645–59.
2. *Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer.* 2015;**136**(5):E359–86.
3. *Marquardt JU, Galle PR, Teufel A.* Molecular diagnosis and therapy of hepatocellular carcinoma (HCC): an emerging field for advanced technologies. *J Hepatol.* 2012;**56**(1):267–75.
4. *Abelev GI, Eraiser TL.* Cellular aspects of alpha-fetoprotein reexpression in tumors. *Semin Cancer Biol.* 1999;**9**(2):95–107.
5. *Behne T, Copur MS.* Biomarkers for hepatocellular carcinoma. *Int J Hepatol.* 2012;**2012**:859076.
6. *Rich N, Singal AG.* Hepatocellular carcinoma tumour markers: current role and expectations. *Best Pract Res Clin Gastroenterol.* 2014;**28**(5):843–53.
7. *Kotelnikova EA, Logacheva MD, Nabieva ER, Pyatnitskiy MA, Vinogradov DV, Makarova AS, Demin AV, Paleeva AG, Kremenetskaya OS, Penin AA, Klepikova AV, Kasianov AS, Shavochkina DA, Kudashkin NE, Patyutko YuI, Mugue NS, Kondrashov AS, Lazarevich NL.* NGS-based identification of druggable alterations and signaling pathways – hepatocellular carcinoma case report. *Biopolym Cell.* 2015; **31**(6):436–46.
8. *Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem.* 2009;**55**(4):611–22.
9. *Fu LY, Jia HL, Dong QZ, Wu JC, Zhao Y, Zhou HJ, Ren N, Ye QH, Qin LX.* Suitable reference genes for real-time PCR in human HBV-related hepatocellular carcinoma with different clinical prognoses. *BMC Cancer.* 2009;**9**:49.
10. *Gao Q, Wang XY, Fan J, Qiu SJ, Zhou J, Shi YH, Xiao YS, Xu Y, Huang XW, Sun J.* Selection of reference genes for real-time PCR in human hepatocellular carcinoma tissues. *J Cancer Res Clin Oncol.* 2008;**134**(9):979–86.
11. *Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, Thorgeirsson SS, Sun Z, Tang ZY, Qin LX, Wang XW.* A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Res.* 2010;**70**(24):10202–12.
12. *Huang Q, Lin B, Liu H, Ma X, Mo F, Yu W, Li L, Li H, Tian T, Wu D, Shen F, Xing J, Chen ZN.* RNA-Seq analyses generate comprehensive transcriptomic landscape and reveal complex transcript patterns in hepatocellular carcinoma. *PLoS One.* 2011;**6**(10):e26168.
13. *Yu K, Ganesan K, Tan LK, Laban M, Wu J, Zhao XD, Li H, Leung CH, Zhu Y, Wei CL, Hooi SC, Miller L, Tan P.* A precisely regulated gene expression cassette potently modulates metastasis and survival in multiple solid cancers. *PLoS Genet.* 2008;**4**(7):e1000129.
14. *Dong H, Zhang L, Qian Z, Zhu X, Zhu G, Chen Y, Xie X, Ye Q, Zang J, Ren Z, Ji Q.* Identification of HBV-MLL4 Integration and Its Molecular Basis in Chinese Hepatocellular Carcinoma. *PLoS One.* 2015;**10**(4):e0123175.
15. *Liu G, Hou G, Li L, Li Y, Zhou W, Liu L.* Potential diagnostic and prognostic marker dimethylglycine dehydrogenase

(DMGDH) suppresses hepatocellular carcinoma metastasis in vitro and in vivo. *Oncotarget.* 2016;**7**(22):32607–16.
16. *Anders S, Huber W.* Differential expression analysis for sequence count data. *Genome Biol.* 2010;**11**(10):R106.
17. *Adnane L, Trail PA, Taylor I, Wilhelm SM.* Sorafenib (BAY 43-9006, Nexavar), a dual-action inhibitor that targets RAF/MEK/ERK pathway in tumor cells and tyrosine kinases VEGFR/PDGFR in tumor vasculature. *Methods Enzymol.* 2006;**407**:597–612.
18. *Makarova AS, Lazarevich NL.* Deregulation of signaling pathways involved in sorafenib resistance of hepatocellular carcinoma. *Klin Lab Diagn.* 2013;(10):66–8.
19. *Shafizadeh N, Ferrell LD, Kakar S.* Utility and limitations of glypican-3 expression for the diagnosis of hepatocellular carcinoma at both ends of the differentiation spectrum. *Mod Pathol.* 2008;**21**(8):1011–8.
20. *Bruix J, Reig M, Sherman M.* Evidence-Based Diagnosis, Staging, and Treatment of Patients With Hepatocellular Carcinoma. *Gastroenterology.* 2016;**150**(4):835–53.
21. *Henry NL, Hayes DF.* Cancer biomarkers. *Mol Oncol.* 2012;**6**(2):140–6.
22. *Marrero JA, Feng Z, Wang Y, Nguyen MH, Befeler AS, Roberts LR, Reddy KR, Harnois D, Llovet JM, Normolle D, Dalhgren J, Chia D, Lok AS, Wagner PD, Srivastava S, Schwartz M.* Alpha-fetoprotein, des-gamma carboxyprothrombin, and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. *Gastroenterology.* 2009;**137**(1):110–8.
23. *Tsuchiya N, Sawada Y, Endo I, Saito K, Uemura Y, Nakatsura T.* Biomarkers for the early diagnosis of hepatocellular carcinoma. *World J Gastroenterol.* 2015;**21**(37):10573–83.
24. *Hass HG, Jobst J, Scheurlen M, Vogel U, Nehls O.* Gene expression analysis for evaluation of potential biomarkers in hepatocellular carcinoma. *Anticancer Res.* 2015;**35**(4):2021–8.
25. *Jia X, Liu J, Gao Y, Huang Y, Du Z.* Diagnosis accuracy of serum glypican-3 in patients with hepatocellular carcinoma: a systematic review with meta-analysis. *Arch Med Res.* 2014;**45**(7):580–8.
26. *Wei T, Zhang LN, Lv Y, Ma XY, Zhi L, Liu C, Ma F, Zhang XF.* Overexpression of platelet-derived growth factor receptor alpha promotes tumor progression and indicates poor prognosis in hepatocellular carcinoma. *Oncotarget.* 2014;**5**(21):10307–17.

Ідентифікація PDGFA як можливого біомаркера, що секретується, гепатоцелюлярної карциноми на підставі транскриптомного аналізу.

М. С. Чесноков, О. М. Кривцова, П. А. Сковороднікова,
А. С. Макарова, И. Ф. Кустова, М. Д. Логачова,
А. А. Пенін, А. В. Клепікова, Д. А. Шавочкіна,
Н. Е. Кудашкін, Е. А. Мороз, Ю. В. Патютко,
Е. А. Котельнікова, Н. Л. Лазаревич

Мета. Для діагностики гепатоцелюлярної карциноми (ГК) ідентифкація потенційних біомаркерів, що секретується. **Методи.** Гени, експресія яких підвищена в тканині ГК і які кодують білки, що секретуються, виявляли РНК-секвенуванням. Експресію генів оцінювали ЗТ-ПЛР або використовували інформацію з відкритих баз даних. Біомаркерні властивості оцінювали за допомогою ROC-кривих, аналізу кореляцій і виживання. **Результати.** РНК-секвенування п'яти випадків ГК виявило гіперекспресію *PDGFA*, що кодує секретуємий білок. Підвищення експресії *PDGFA* та *GPC3* виявлено в 17 з 19 зразків ГК та у більшості випадків з баз даних. Комбінація з *PDGFA* та *GPC3* розрізнює тканину ГК і непухлинну тканину печінки краще, ніж *PDGFA* або *GPC3* окремо. Гіперекспресія *PDGFA* асоційована з кращим прогнозом для пацієнтів з ранніми стадіями ГК і низькою інвазією пухлини в судини. **Висновки.** PDGFA – перспективний біомаркер ГК, що секретується, який може бути використаний разом з GPC3 для підвищення його чутливості.

Keywords: hepatocellular carcinoma; PDGFA; tumor biomarker; NGS

Идентификация PDGFA как возможного секретируемого биомаркера гепатоцеллюлярной карциномы на основании транскриптомного анализа.

М. С. Чесноков, О. М. Кривцова, П. А. Сковородникова,
А. С. Макарова, И. Ф. Кустова, М. Д. Логачева,
А. А. Пенин, А. В. Клепикова, Д. А. Шавочкина,
Н. Е. Кудашкин, Е. А. Мороз, Ю. И. Патютко,
Е. А. Котельникова, Н. Л. Лазаревич

Цель. Поиск потенциальных секретируемых биомаркеров для диагностики гепатоцеллюлярной карциномы (ГК). **Методы.** Гены, экспрессия которых повышена в ткани ГК и кодирующие секретируемые белки, выявляли РНК-секвенированием. Экспрессию генов оценивали ОТ-ПЦР или использовали информацию из открытых баз данных. Биомаркерные свойства оценивали с помощью ROC-кривых, анализа корреляций и выживаемости. **Результаты.** РНК-секвенирование 5 случаев ГК выявило гиперэкспрессию *PDGFA*, кодирующего секретируемый белок. Повышение экспрессии *PDGFA* и *GPC3* выявлено в 17 из 19 образцов ГК и в большинстве случаев из баз данных. Совместное использование двух маркерных генов *PDGFA* и *GPC3* позволяет дифференцировать ткань ГК от неопухолевой ткани печени лучше, чем *PDGFA* или *GPC3* по отдельности. Гиперэкспрессия *PDGFA* ассоциирована с лучшим прогнозом для пациентов с ранними стадиями ГК и низкой инвазией опухоли в сосуды. **Выводы.** PDGFA – перспективный секретируемый биомаркер ГК, который может быть использован вместе с GPC3 для повышения его чувствительности.

Keywords: hepatocellular carcinoma; PDGFA; tumor biomarker; NGS

Received 01.11.2016