EUROPEAN JOURNAL OF CANCER 44 (2008) 885–897

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Molecular mapping of human hepatocellular carcinoma provides deeper biological insight from genomic data

Nobuyoshi Kittaka[a], Ichiro Takemasa[a,], Yutaka Takeda[a], Shigeru Marubashi[a], Hiroaki Nagano[a], Koji Umeshita[a], Keizo Dono[a], Kenichi Matsubara[b], Nariaki Matsuura[c], Morito Monden[a]*

[a]*Department of Surgery, Graduate School of Medicine, Osaka University, 2-2 Yamadaoka E-2, Suita, Osaka 565-0871, Japan*
[b]*DNA Chip Research Inc., 1-1-43 Suehirocho, Tsurumi-ku, Yokohama 230-0045, Japan*
[c]*Department of Functional Diagnostic Science, Osaka University Graduate School of Medicine, 1-7 Yamadaoka, Suita, Osaka 565-0871, Japan*

ARTICLE INFO

Article history:
Received 5 December 2007
Received in revised form
5 February 2008
Accepted 12 February 2008
Available online 11 March 2008

Keywords:
DNA microarray
Network analysis
Integrative method
'Hotspot' region
Biological insight

ABSTRACT

DNA microarray analysis of human cancer has resulted in considerable accumulation of global gene profiles. However, extraction and understanding the underlying biology of cancer progression remains a significant challenge. This study applied a novel integrative computational and analytical approach to this challenge in human hepatocellular carcinoma (HCC) with the aim of identifying potential molecular markers or novel therapeutic targets. We analysed 100 HCC tissue samples by human 30 K DNA microarray. The gene expression data were uploaded into the network analysis tool, and the biological networks were displayed graphically. We identified several activated 'hotspot' regions harbouring a concentration of upregulated genes. Several 'hotspot' regions revealed integrin and Akt/NF-κB signalling. We identified key members linked to these signalling pathways including osteopontin (SPP1), glypican-3 (GPC3), annexin 2 (ANXA2), S100A10 and vimentin (VIM). Our integrative approach should significantly enhance the power of microarray data in identifying novel potential targets in human cancer.

1. Introduction

Investigation of various cancers at the molecular level is well underway through functional approaches including DNA microarray technology that can simultaneously detect the expression levels of tens of thousands of genes. The resulting wealth of data has been analysed with a variety of clustering, partitioning and pattern-matching algorithms in the quest to generate molecular signatures for several human malignant tumours with respect to their stage, prognostic outcome and response to therapy.

Notwithstanding the obvious power of the genomic data generated, these molecular analyses have not yielded the expected advances in our understanding of the mechanisms of cancer development, or the identification of critical genomic and molecular aberrations that would improve the precision of diagnosis or serve as therapeutic targets. This is mainly due to the overwhelming diversity of genome-wide interactions and gene-expression patterns, which limit effective learning from experimental data alone. Network analysis technologies are currently addressing this problem by mapping the gene expression data into relevant networks based on known mammalian biology, derived from basic and clinical research. To this end, our group has combined microarray analysis with a computational tool to obtain further biological insights into the regulatory networks of differentially expressed genes and

* *Corresponding author:* Tel.: +81 6 6879 3251; fax: +81 6 6879 3259.
E-mail address: alfa-t@sf6.so-net.ne.jp (I. Takemasa).
0959-8049/$ - see front matter
doi:10.1016/j.ejca.2008.02.019

the corresponding canonical pathways related to the progression of cancer. We applied this integrative approach to human hepatocellular carcinoma (HCC), the fifth most common malignancy worldwide.[1,2] Despite the remarkable improvements in diagnosis and patient management, the outcome for patients with HCC remains grave, mainly due to the advanced tumour stage accelerated by intrahepatic tumour spread and frequent tumour recurrence.[3] Hepatocarcinogenesis is a multistep process involving somatic mutations, loss of tumour suppressor genes and possibly the activation or overexpression of certain oncogenes.[4] These events lead to changes in the expression of numerous genes, and comparison of gene expression patterns between HCC and normal liver tissue is a popular method for characterising tumour properties and identifying novel target genes for possible therapy. However, this method has not proven to be sufficiently definitive in identifying genetic determinants of specific HCC regulatory pathways. New approaches are urgently needed to better understand the underlying mechanisms of hepatocarcinogenesis, and to develop new therapeutic approaches targeted to HCC-specific molecular abnormalities. By highlighting several activated regions in the genome (known as 'hotspot' regions[5,6]) involved in regulating the progression of HCC, we have identified significantly upregulated genes linked to these 'hotspot' pathways as potential key molecules.

Our integrative analysis revealed two 'hotspot' canonical pathways (integrin and Akt/NF-κB signalling pathways) and identified five potential key genes that were upregulated in the majority of HCC tumours. We further investigated two of these potential key molecules, ANXA2 and S100A10, which were upregulated at the protein and mRNA levels in most HCC samples. Importantly, because it is proteins that function in networks controlling critical cellular events,[7] it is reasonable to speculate that coexpression of ANXA2 and S100A10 at the protein level might have an impact on hepatocarcinogenesis through the activated 'hotspot' pathway.

2. Materials and methods

2.1. *Tissue samples*

Samples from 100 HCC tissues and seven normal livers without virus infection were obtained with informed consent from patients who underwent hepatic resection at Osaka University Hospital from 1997 to 2003. Tissue specimens (approximately 5 mm^3) for RNA isolation were stored at –80 °C until use. All tissue specimens were submitted for routine pathological evaluation and confirmation of diagnosis. The histopathological characterisation of HCC was based on the Classification of the Liver Cancer Study Group of Japan. Table 1 lists the clinicopathological features of the 100 cases of HCC.

2.2. *Extraction and quality assessment of RNA*

Total RNA was purified from tissue samples using TRIzol reagent (Invitrogen, San Diego, CA) as described by the manufacturer. The integrity of RNA was assessed on an Agilent 2100 Bioanalyzer and RNA 6000 LabChip kits (Yokokawa Ana-

Table 1 – Clinicopathological characteristics of 100 patients with HCC

Clinicopathological features	*n*
Age	
Median	66
Range	47–81
Gender	
Male	81
Female	19
Virus	
HBV	21
HCV	40
Both	28
None	11
Child-Turcotte-Pugh stage	
A	77
B	23
C	0
Liver cirrhosis	
Present	42
Absent	58
AFP	
<200 ng/ml	71
⩾200 ng/mg	29
PIVKA-II	
<50mAU/ml	36
⩾50mAU/ml	64
Tumour size	
<5.0 cm	76
⩾5.0 cm	24
Edmonson grading	
1–2	43
3–4	57
Histologic type of tumour	
Well differentiated	4
Moderately differentiated	41
Poorly differentiated	55
Vascular invasion	
Present	41
Absent	59
Intrahepatic metastasis	
Present	22
Absent	78
Pathological stage	
I	23
II	52
III	20
IVA	5
CLIP score	
0	56
1	35
2	8
3	0
4	1
5	0
6	0
JIS score	
0	18
1	46
2	26
3	9
4	1
5	0

CLIP score; The cancer of Liver Italian Program score.
JIS score; The Japan Integrated Staging score.

lytical Systems, Tokyo, Japan). Only high-quality RNA with intact 18s and 28s RNA was used for subsequent analysis. Seven RNA extractions from different normal liver tissue were mixed as the control reference.

2.3. Preparation of fluorescently labelled aRNA targets and hybridisation

Extracted RNA samples were amplified with T7 RNA polymerase using the Amino Allyl MessageAmp™ aRNA kit (Ambion, Austin, TM) according to the protocol provided by the manufacturer. The quality of each Amino Allyl-aRNA sample was checked on the Agilent 2100 Bioanalyzer. Five micrograms of control and experimental aRNA samples were labelled with Cy3 and Cy5, respectively, mixed and hybridised on an oligonucleotide microarray covering 30,336 human probes (AceGene Human 30K; DNA Chip Research Inc. and Hitachi Software Engineering Co., Yokohama, Japan). The experimental protocol is available at http://www.dna-chip.co.jp/thesis/AceGeneProtocol.pdf. The microarrays were scanned on a ScanArray 4000 (GSI Lumonics, Billerica, MA).

2.4. Analysis of microarray data

Signal values were calculated using DNASIS Array Software (Hitachi Software Inc., Tokyo). Following background subtraction, data with low signal intensities were excluded from additional investigation. In each sample, the Cy5/Cy3 ratio values were log-transformed. Then, global equalisation to remove a deviation of the signal intensity between whole Cy3- and Cy5-fluorescence was performed by subtracting the median of all log(Cy5/Cy3) values from each log(Cy5/Cy3) value. Genes with missing values in more than 20% of samples were excluded from further analysis; a total of 16,923 genes out of 30,336 were available for analysis.

2.5. Gene network analysis

We further analysed the signature genes of HCC by Ingenuity Pathways Analysis (Ingenuity systems, Mountain View, CA; http://www.ingenuity.com), a web-delivered application that enables biologists to discover, visualise and explore relevant networks significant to their experimental results, such as gene expression array datasets. The application makes use of the Ingenuity Pathways Knowledge Base (IPKB), which contains large amounts of individually modelled relationships between gene objects (e.g., genes, mRNAs and proteins) to dynamically generate significant biological networks and pathways. The submitted genes that are mapped to the corresponding gene objects in the IPKB are called 'focus genes'. The focus genes are used as the starting point for generating biological networks. To start building a network, the Ingenuity software queries the IPKB for interactions between focus genes and all the other genes stored in the IPKB, and then generates a set of networks with a maximum network size of 35 genes. A p value for each network is calculated according to the fit of the user's set of significant genes. This is accomplished by comparing the number of focus genes that participate in a given network relative to the total number of occurrences of those genes in all networks stored in the IPKB. The score of a network is displayed as the negative log of the p value, indicating the probability that a collection of genes equal to or greater than the number in a network could be achieved by chance alone.

2.6. Selection of candidate genes expressed in HCC

To identify molecular pathways that may be activated or suppressed in HCC, we used a network knowledge-base approach, Ingenuity Pathway Analysis Software, to analyse genome-wide transcriptional responses in the context of known functional interrelationships amongst proteins, small molecules and phenotypes. The post-normalised genes (16,923 genes) either up- or down-regulated in the microarray data, were uploaded into the IPKB as a tab-delimited text file of GenBank accession numbers. These biological networks comprising 5936 genes are displayed graphically as nodes (genes/gene products) and edges (the biological relationships between the nodes). The nodes are displayed using various shapes that represent the functional class of the gene product. The colour green reflects downregulation of gene expression, and red represents upregulation of gene expression with the significance of that regulation represented by colour intensity. Edges are displayed with various labels that describe the nature of the relationship between the nodes. In this way, simultaneous survey and evaluation of the subnetwork regions enabled us to identify several activated canonical pathways in HCC. We highlighted new molecules linked to the 'hotspot' canonical pathways.

2.7. Real-time quantitative RT-PCR analysis

Total RNA (1 μg) was used for reverse transcription, and complementary DNA (cDNA) was generated using the Reverse Transcription System (Promega, Madison, WI) as described previously.[8] Quantification of mRNA expression of the candidate genes listed in Table 2 was performed using a real-time thermal cycler, the LightCycler and detection system (Roche Diagnostics, Mannheim, Germany). For detection of the amplification products, LightCycler-DNA master SYBR green I (Boehringer

Table 2 – Candidate genes and expression ratio of microarray analysis

CDS ID	Gene symbol	Description	Average of Cy5/Cy3
NM_000582	SPP1	Secreted phosphoprotein 1 (osteopontin)	4.69
NM_004484	GPC3	Glypican 3	4.23
NM_004039	ANXA2	Annexin 2	2.86
M38591	S100A10	Cellular ligand of annexin 2	1.97
NM_003380	VIM	Vimentin	1.82

Mannheim, Mannheim, Germany) was used as described previously.[9] Briefly, a 20 μl reaction volume containing 2 μl of cDNA and 0.2 μmol/l of each primer was applied to a glass capillary. The primer sequences, PCR cycle conditions and annealing temperatures are listed in Supplementary Table 1. Quantitative analysis of mRNA was performed using LightCycler analysis software (Roche Diagnostics). The relative expression level of the candidate gene was computed with respect to the internal standard GAPDH mRNA to normalise for variations in the amount of input cDNA. The level of expression of the candidate gene was provided by the ratio, in which each normalised gene value in tissue samples was divided by GAPDH mRNA in the same control reference used in the microarray assay. We compared the ratio of candidate genes between samples randomly selected out of 100 HCC samples.

2.8. Immunohistochemical staining

Formalin-fixed, paraffin-embedded samples were cut into 5 μm sections, and these were deparaffinised in xylene and rehydrated through a graded series of ethanol. Immunohistochemical staining was performed using a Vectastain ABC peroxidase kit (Vector Labs, Burlingame, CA) as described previously.[10] Briefly, the sections were treated for antigen retrieval in 0.01 M sodium citrate buffer (pH 6.0) for 40 min at 95 °C, followed by incubation in methanol containing 0.3% hydrogen peroxide at room temperature for 20 min to block endogenous peroxidase. After blocking endogenous biotin, the sections were incubated with normal protein-block serum solution at room temperature for 20 min, to block non-specific staining, and then incubated overnight at 4 °C with anti-ANXA2 (mouse monoclonal IgG, diluted 1:500, Abcam Inc.), anti-S100A10 (mouse monoclonal IgG, diluted 1:400, Swant Inc.) and anti-GPC3 (mouse monoclonal IgG, University of Toronto, Jorge Filmus et al.[11]) as primary antibodies. After washing three times for 5 min in phosphate buffered saline (PBS), sections were incubated with a biotin-conjugated secondary antibody (horse anti-mouse for ANXA2, S100A10 and GPC3) at room temperature for 20 min and finally incubated with peroxidase-conjugated streptavidin at room temperature for 20 min. The peroxidase reaction was then developed with 3,3′-diaminobenzidine tetrachloride (Wako Pure Chemical Industries, Osaka, Japan). Finally, the sections were counterstained with Mayer's haematoxylin. For negative controls, sections were treated the same way except they were incubated with non-immunised rabbit IgG or Tris-buffered saline instead of the primary antibody. Immunohistochemical staining was assessed by two investigators independently, without the knowledge of the corresponding clinicopathological data.

2.9. Statistical analysis

Pearson's correlation coefficient, χ^2 test, *t*-test and Kaplan–Meier plot were analysed using StatView (Version 5.0) software (Abacus Concepts, Berkeley, CA). p values less than 0.05 were considered statistically significant. Hierarchical cluster analysis (HCA) was performed with Euclidean distance coefficient as a similarity coefficient and the unweighted pair group method using arithmetic averages (UPGMA) as the clustering algorithm, using GeneMaths (Version 2.0) software.

3. Results

3.1. Microarray analysis of gene expression changes in HCC tumours

Gene expression profiling of primary HCC tumours from 100 patients was examined by DNA microarray. We calculated the mean expression levels of each gene across all HCC samples, and, as a preliminary analysis, identified the top 2% of candidate genes displaying at least a 1.5-fold increase in expression. These highly upregulated genes included α-fetoprotein (AFP; data not shown), a common prognostic marker for HCC (fold change = 1.56), and GPC3, recently identified as a novel tumour marker of HCC (fold change = 4.23; fourth highest upregulation).

3.2. Identification of biologically relevant networks and potential key genes highly expressed in HCC tumours

In our global network comprising 5936 genes (Supplementary Fig. 1), we highlighted integrin and Akt/NF-κB signalling as two 'hotspot' pathways that comprised a concentration of upregulated genes. The integrin signalling pathway shown in Fig. 1A (gene subnetworks listed in Supplementary Table 2) was identified as significantly activated in HCC and contained 11 upregulated genes, flagging this pathway as a key regulator in HCC tumourigenesis. This fits with the role of integrin signalling in promoting cell proliferation and cell motility.[12] Furthermore, SPP1 and GPC3 were identified as potential key genes (upregulated with >1.5-fold change), with links to integrin signalling. Similarly, we identified the activation of the Akt/NF-κB pathway shown in Fig. 1B in HCC tumours (gene subnetworks listed in Supplementary Table 3). This signalling pathway contained 12 upregulated genes and plays key roles in many cell processes relevant to tumourigenesis including cell survival and apoptosis.[13] Amongst the genes linked to Akt/NF-κB signalling and that had >1.5-fold change were ANXA2, S100A10 and VIM. Network analysis revealed a link between Akt/NF-κB signalling and both ANXA2 and S100A10 through interactions with β-actin (ACTB) and E-cadherin (CDH1). These candidate genes are listed in Table 2.

3.3. Quantitative RT-PCR validation of several selected genes

To validate the microarray data, we performed quantitative RT-PCR for candidate genes in 20 samples randomly selected out of the 100 HCC tissues. We compared gene expression levels generated from quantitative RT-PCR with those from microarray analysis by Pearson's correlation coefficients for each candidate gene and StatView Software (Fig. 2). Each of the five genes analysed showed significant correlation confirming the results obtained by DNA microarray.

3.4. Immunohistochemical study of glypican 3, Annexin 2 and S100A10 in patient samples

Of the five genes overexpressed in HCC tumours by RT-PCR, immunohistochemical staining for GPC3 was performed on 10 samples of HCC and surrounding non-cancerous tissues

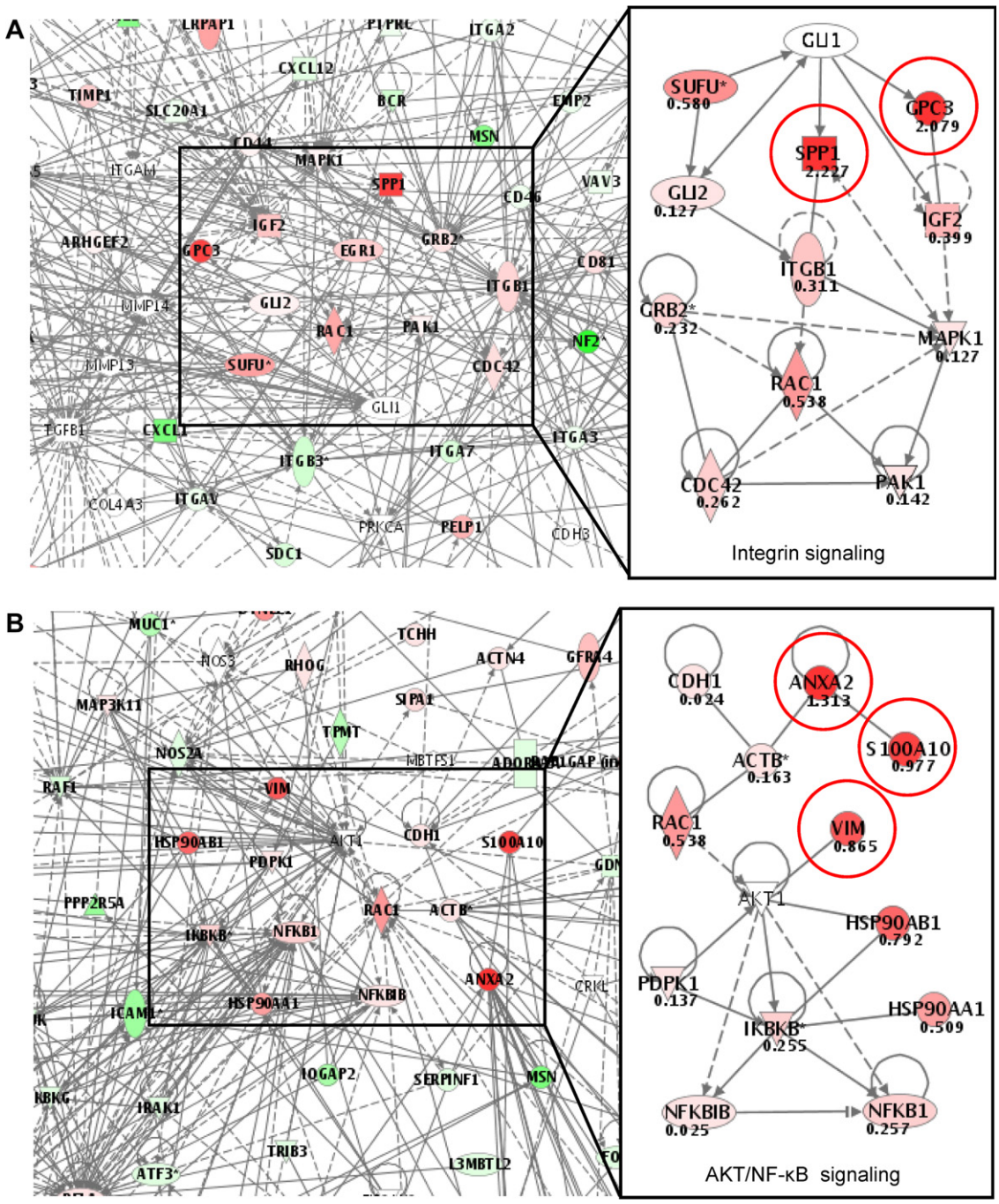

Fig. 1 – (A) The integrated method of DNA microarray and Ingenuity Pathway Analysis produced a network with 'hotspot' regions harbouring concentrations of upregulated genes. These genes included ITGB1 (integrin, beta 1), RAC1 (ras-related C3 botulinum toxin substrate 1), GRB2 (growth factor receptor-bound protein 2), CDC42 (cell division cycle 42), PAK1 (p21/Cdc42/Rac1-activated kinase 1) and MAPK1, which are all associated with integrin signalling. SPP1, GPC3, GLI1, GLI2, SUFU (suppressor of fused homolog) and IGF2 are also linked to this pathway. The circled genes, SPP1 and GPC3, were selected as candidate genes. The numerical value of each gene represents the median of all log(Cy5/Cy3) values. (B) The integrative method showed a second network including a 'hotspot' region. This region contained AKT1 (v-akt murine thymoma viral oncogene homolog 1), PDPK1 (3-phosphoinositide-dependent protein kinase-1), NFKBIB (nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, beta), NFKB1 (nuclear factor of kappa light polypeptide gene enhancer in B-cells 1), IKBKB (inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta), HSP90AA1 (heat shock protein 90 kDa alpha, class A member 1) and HSP90AB1 (heat shock protein 90 kDa alpha, class B member 1), which are all associated with the Akt/NF-κB signalling pathway. ANXA2, S100A10, ACTB, CDH1 and VIM are also linked to this pathway. The circled genes, ANXA2, S100A10 and VIM, were selected as candidate genes. The numerical value of each gene represents the average of all log(Cy5/Cy3) values.

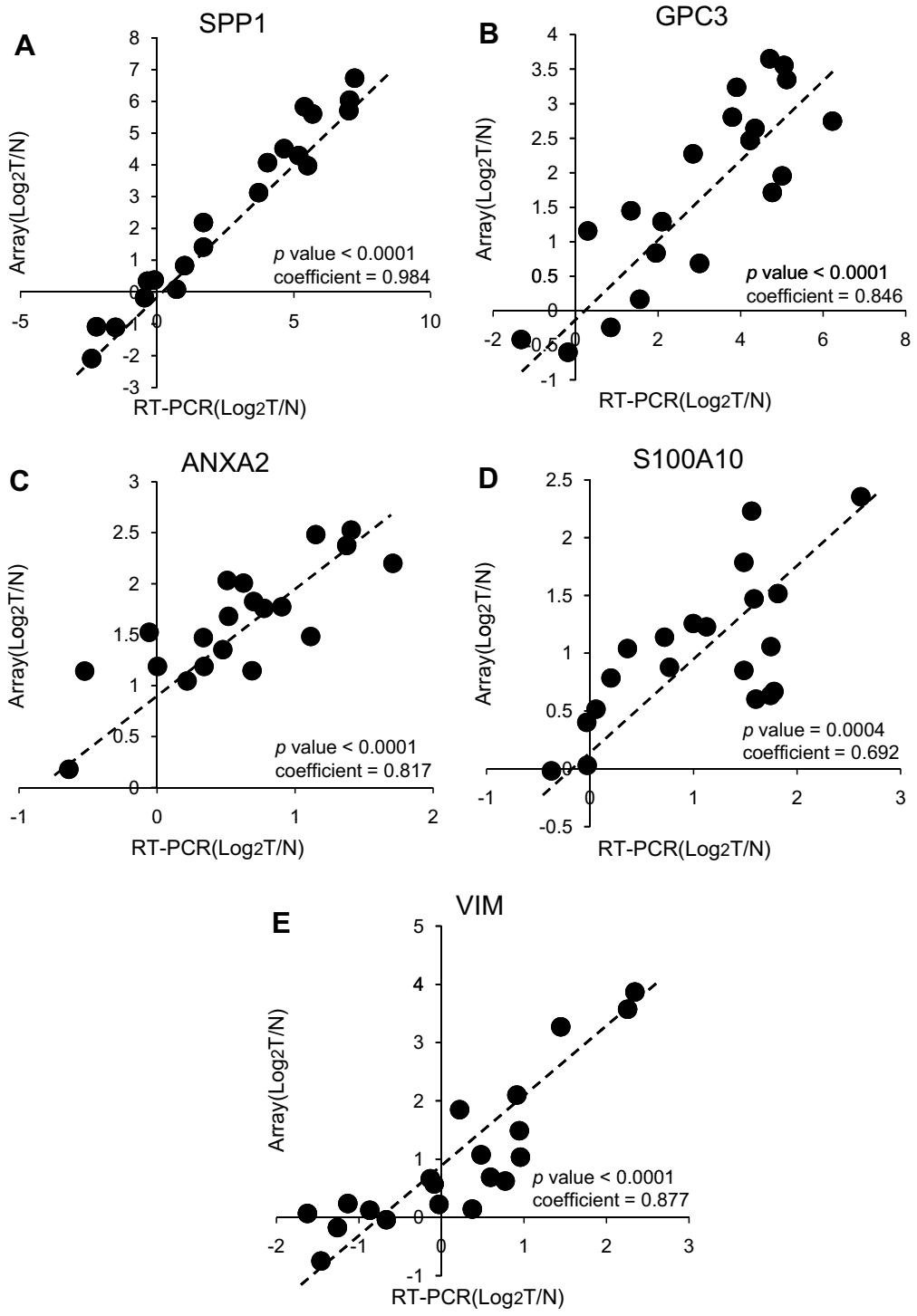

Fig. 2 – Results of quantitative RT-PCR on 20 samples randomly selected from the 100 HCC samples. Individual mRNA levels were normalised to GAPDH and expressed relative to those in a mixture of seven normal livers for SPP1, GPC3, ANXA2, S100A10 and VIM. We compared gene expression levels generated from quantitative RT-PCR with those from microarray analysis and used Pearson's correlation analysis for each candidate gene using StatView Software. (A), (B), (C), (D) and (E) show the correlation of SPP1, GPC3, ANXA2, S100A10 and VIM, respectively.

(Fig. 3A and B). GPC3 expression was observed in 7 of 10 cases of moderately or poorly differentiated HCC. As published previously,[11] staining of GPC3 was observed in a coarsely granular pattern near the cell membrane (2/7) and dispersed evenly in the cytoplasm (5/7). GPC3 expression was undetectable in all non-neoplastic tissues with diffuse hepatitis changes.

Immunohistochemical staining of ANXA2 and S100A10 was then performed on 20 paraffin-embedded samples of HCC and surrounding non-cancerous tissues. The Ca^{2+}-

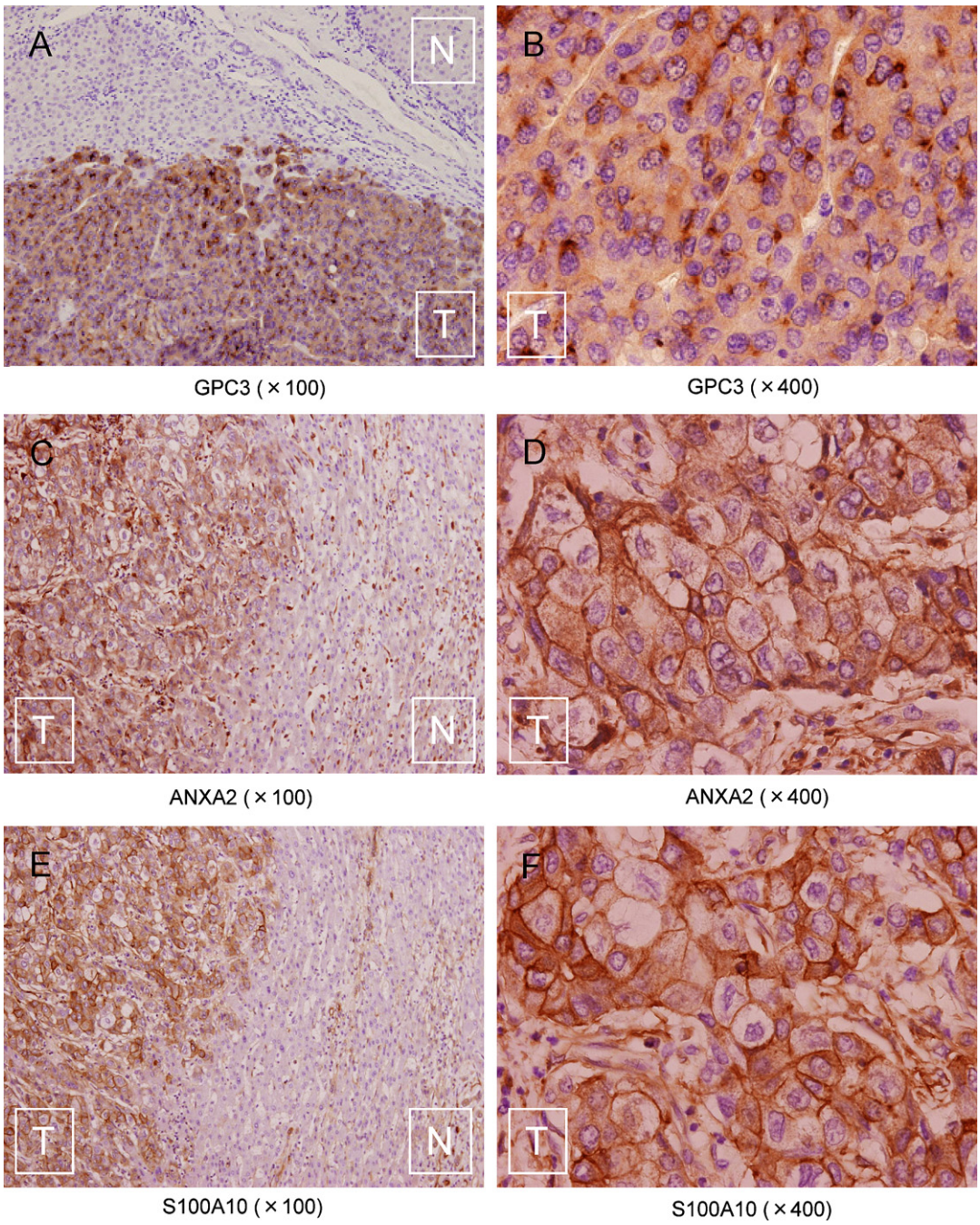

Fig. 3 – (A and B) GPC3 protein expression in human HCC tissue. Brown GPC3 immunostaining is evident in cancer cells. Note the diffuse non-granular staining pattern in the cytoplasm. (C and D) ANXA2 protein expression in human HCC tissue. Note ANXA2 staining in the cell membrane of cancer cells, with slight immunoreactivity in the cytoplasm. (E and F) S100A10 protein expression in human HCC tissues. Note S100A10 staining at the cell membrane of cancer cells. C, D, E and F are from the same tissue sample, and staining for S100A10 overlapped significant with ANXA2 staining. (A, C and E) ×100 magnification. (B, D and F) ×400 magnification. T; tumour region. N; normal liver.

and membrane-binding protein ANXA2 can form a hetero-tetrameric complex with S100A10 and this complex is thought to serve as a bridging or scaffolding function in the membrane underlying cytoskeleton.[14] Previous studies demonstrated both ANXA2 and S100A10 at the plasma membrane in hepatoblastoma HepG2 cell lines,[15] but not in human HCC tissues. Immunohistochemical staining of ANXA2 and S100A10 was stronger at the plasma membrane of the same samples than in the cytoplasm (Fig. 3C–F). The immunoreactivity for ANXA2 was heterogeneous, and cancerous tissues were immunopositive in 16 samples. Endothelial cells were immunopositive for ANXA2 in all samples tested. Similarly, staining of S100A10 was heterogeneous, and cancerous tissues were immunopositive in 17 samples.

Table 3 – Clinicopathological characteristics and results of immunohistochemical staining of ANXA2 and S100A10

Patient	Age	Gender	Hepatitis virus infection	Histological type	Vascular invasion	Annexin2 expression	S100A10 expression
Case 1	71	M	HCV	por	–	++	++
Case 2	66	F	HBV	mod	+	++	++
Case 3	38	M	HBV	mod	+	++	++
Case 4	65	M	HCV	por	+	++	++
Case 5	59	M	HCV	por	–	++	++
Case 6	72	F	HCV	mod	–	+	++
Case 7	71	M	–	por	–	++	+
Case 8	72	M	HBV, HCV	por	–	++	+
Case 9	61	M	–	por	–	++	+
Case 10	78	M	HCV	por	–	++	+
Case 11	49	F	HBV	por	+	++	+
Case 12	72	F	HBV	por	+	+	+
Case 13	68	M	HBV, HCV	mod	–	+	+
Case 14	78	F	HBV, HCV	mod	–	+	+
Case 15	74	M	HBV, HCV	well	–	+	+
Case 16	67	M	–	mod	–	–	+
Case 17	53	M	HBV	mod	+	–	–
Case 18	75	M	HCV	por	–	–	–
Case 19	66	M	HBV, HCV	mod	+	–	–
Case 20	60	M	HCV	por	–	–	–

++, strong immunopositive; +, partial immunopositive; –, immunonegative.

Colocalisation of ANXA2 and S100A10 was observed in 15 samples (Table 3).

3.5. *Correlation between gene expression signature of the two 'hotspot' and clinicopathological features*

Next, to better understand if any of the two 'hotspot' identified by this integrated approach correlates with clinicopathological features, a hierarchical clustering of all 100 HCC samples using the upregulated genes included in the 'hotspot' was performed. Fig. 4A shows the gene expression profiles using the 11 genes upregulated in the 'hotspot 1 (integrin signalling)'. Examination of this result allowed identification of three subgroups; 'relatively high-activated group ($n = 39$), defined as Group A1', 'intermediate-activated group ($n = 45$), defined as Group A2', and 'relatively low-activated group ($n = 16$), defined as Group A3'. Likewise, Fig. 4B shows the gene expression profiles using the 12 genes upregulated in the 'hotspot 2 (Akt/NF-κB signalling)'. Examination of this result also identified two subgroups; 'relatively high-activated group ($n = 17$), defined as Group B1' and 'relatively low-activated group ($n = 83$), defined as Group B2'. Having identified the two distinctive subgroups; 'relatively high-activated group' and 'relatively low-activated group' in each of the two 'hotspots', we examined the association between the activation of 'hotspot' and clinicopathological data (Table 4A and B). In the integrin signalling, the activated profile was significantly associated with intrahepatic metastasis ($p = 0.012$), tumour size ($p = 0.023$) and Edmonson grading ($p < 0.001$). On the other hand, in the Akt/NF-κB signalling, the activated profile was significantly associated with Edmonson grading ($p = 0.004$). Kaplan–Meier plot showed a significant difference in the probability of disease-free survival ($p = 0.037$) and overall survival ($p = 0.045$) between 'Group A1' and 'Group A3' in the integrin signalling (Fig. 4C and D). In the Akt/NF-κB signalling, a significant difference was observed in disease-free survival ($p = 0.045$, Fig. 4E and F).

3.6. *Overview of the distribution of differentially expressed genes on human chromosome*

To compare our microarray data with chromosomal aberrations in HCC, we also investigated the chromosomal region in which the differentially expressed genes were harboured. The public source for annotating the location of each gene was the National Center for Biotechnology Information (NCBI). This investigation revealed that the regions of high density of 100 upregulated genes tended to be at chromosomes 1q, 6p and 8q (Fig. 5A), whilst those of 100 downregulated genes were at chromosomes 4q and 16q (Fig. 5B). Furthermore, SPP1, GPC3, ANXA2 and S100A10, identified as key molecules, were separately located at chromosomes 4q, Xq, 15q and 1q.

4. Discussion

In the post-genomic period, DNA microarray technology is used to monitor disease progress and to individualise treatment regimens. However, extracting new biological insights from high-throughput genomic studies of cancer progression poses a challenge due to difficulties in recognising and evaluating relevant biological processes from vast quantities of experimental data. Although other high-throughput technologies in protein expression (proteomics) and low-molecular weight metabolite expression (metabolomics) have made remarkable progress, no comprehensive analytical techniques exist that can measure more than 500,000 protein forms and 100,000–1,000,000 metabolites quantitatively.[16] Generating biological networks from comprehensive gene expression profiles manually in a visual manner could be used to navigate

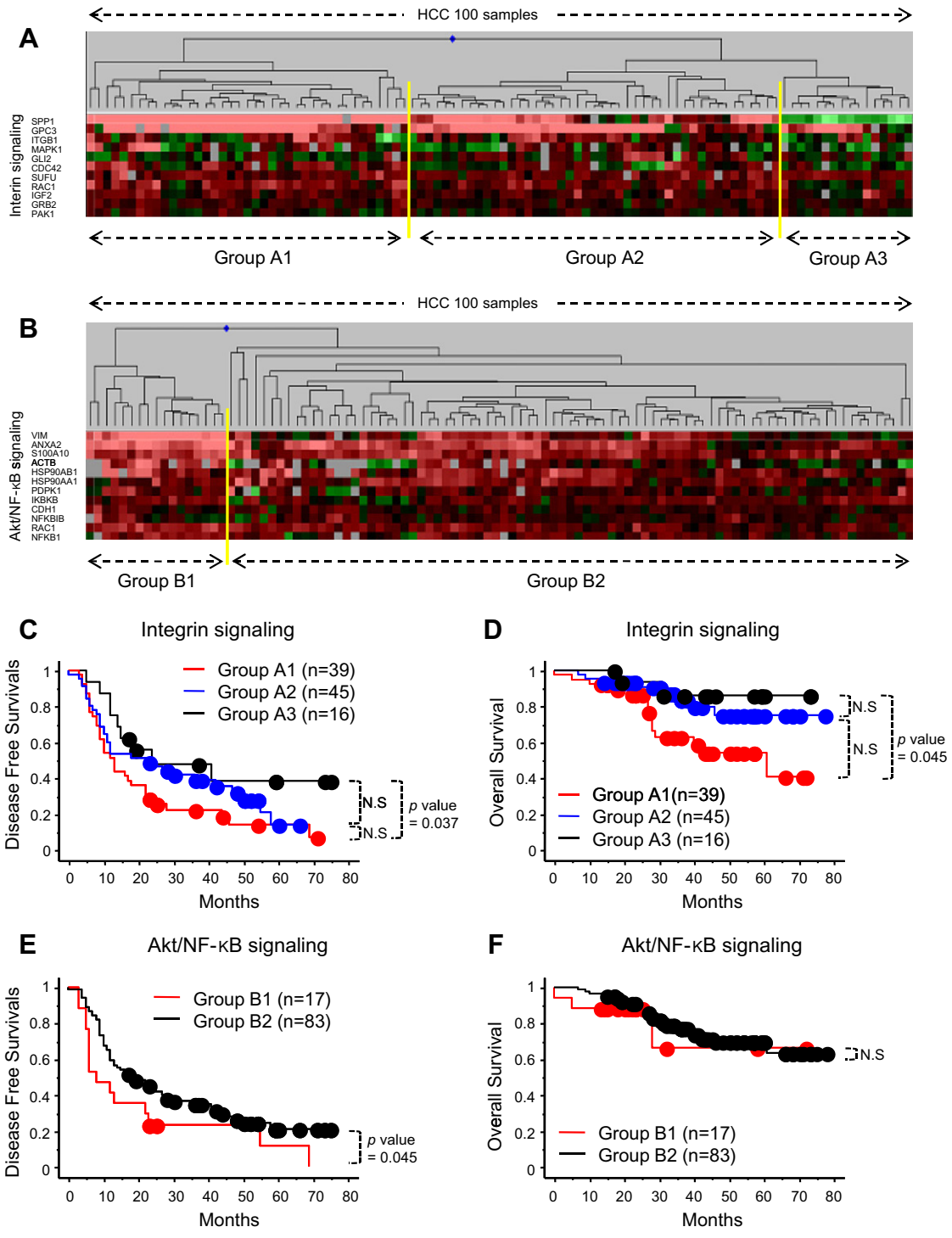

Fig. 4 – (A) Hierarchical clustering analysis of all 100 HCC samples using the 11 upregulated genes included in the 'hotspot 1 (integrin signalling)'. Red and green indicate relative high- and low-expression, respectively. Based on the similarities of their gene expression profiles, samples were grouped in 'relatively high-activated group (n = 39), defined as Group A1', 'intermediate-activated group (n = 45), defined as Group A2', and 'relatively low-activated group (n = 16), defined as Group A3'. (B) A hierarchical clustering analysis of all 100 HCC samples using the 12 upregulated genes included in the 'hotspot 2 (Akt/NF-κB signalling)'. Based on the similarities of their gene expression profiles, samples were grouped in 'relatively high-activated group (n = 17), defined as Group B1' and 'relatively low-activated group (n = 83), defined as Group B2'. (C and D) Disease-free survival and overall survival of each of the activated groups in the 'hotspot 1 (integrin signalling)' (Kaplan–Meier plot). The log-rank p value is shown. NS, not significant. (E and F) Disease-free survival and overall survival of each of the activated groups in the 'hotspot 2 (Akt/NF-κB signalling)' (Kaplan–Meier plot). The log-rank p value is shown. NS, not significant.

Table 4 – Clinical and pathological characteristics of the high-activated and low-activated groups in each of integrin signalling and AKT/NF-κB signalling

Characteristics	Integrin signalling ('Hotspot' 1)			*p* value
	Group A1 (*n* = 39)	Group A2 (*n* = 45)	Group A3 (*n* = 16)	
	No. of patients (%)	No. of patients (%)	No. of patients(%)	
A				
Intrahepatic metastasis	12(30.8)	(10(22.2))	0(0)	0.012
Tumour size (cm)	4.79 ± 3.12	(3.49 ± 1.75)	2.91 ± 1.22	0.023
Edmonson grading				
1–2	11(28.2)	22(48.9)	10(62.5)	<0.001
3–4	28(71.8)	23(51.1)	6(37.5)	
Pathological stage				
I	5(12.8)	15(33.3)	3(18.7)	0.642
II	22(56.4)	20(44.4)	10(62.5)	
III	9(23.1)	8(17.8)	3(18.7)	
IVA	3(7.7)	2(4.5)	0	

Characteristics	Akt/NF-κB signalling ('Hotspot' 2)		*p* value
	Group B1 (*n* = 17)	Group B2 (*n* = 83)	
	No. of patients (%)	No. of patients (%)	
B			
Intrahepatic metastasis	4(23.5)	18(21.7)	0.867
Tumour size (cm)	4.55 ± 2.58	3.88 ± 2.81	0.226
Edmonson grading			
1–2	2(11.7)	41(49.4)	0.004
3–4	15(88.3)	42(50.6)	
Pathological stage			
I	1(5.9)	22(26.5)	0.333
II	11(64.7)	41(49.4)	
III	4(23.5)	16(19.3)	
IVA	1(5.9)	4(4.8)	

p values of *A* are obtained by comparing Group A1 with Group A3.
p values of *B* are obtained by comparing Group B1 with Group B2.
p values for intrahepatic metastasis, Edmonson grading and pathological stage are obtained by χ^2 test.
p value for tumour size is obtained by *t* test.
Tumour size is mean ± SD.

through and unravel the complex networks involved in cancer progression. Here, we combined genome-wide expression analysis with a new bioinformatics method, Ingenuity Pathway Analysis, to clarify the relationship between the microarray datasets and the canonical pathways based on the published literature and identify functional networks, 'hotspot', responsible for the progression of HCC. Furthermore, we discovered several molecules commonly upregulated in HCC as potential key players in the neoplastic process.

This combined approach revealed that several distinct regions with upregulated genes were concentrated. These concentrations of activated genes included several genes involved in the WNT signalling pathway, which has been the subject of intense research in recent years. In addition, we highlighted integrin and Akt/NF-κB as two 'hotspot' signalling pathways, and propose that these signalling pathways are crucial for the core biological functions in HCC progression and potential intervention, such as cell proliferation, cell survival and apoptosis.[17]

In addition to studies of gene expression at the transcriptional level, protein analysis is vital for understanding the regulatory processes in living organisms, because emerging evidence suggests that mRNA expression patterns themselves are necessary but insufficient for quantitative description of biological systems. So far, comparative studies of mRNA and protein abundance indicate that only 20–28% of the total variation of protein abundance can be attributed to mRNA abundance alone.[18] The limiting factors were explained partly by translational processes (microRNAs repress the translation of mRNAs into proteins) and post-translational modification (such as phosphorylation, methylation, acetylation, glycosylation and ubiquitination). In fact, this was the basis for investigating the expression levels of proteins encoded by highly upregulated genes related to key signalling pathways in hepatocarcinogenesis.

Using our analysis protocol, integrin pathway-associated molecules, SPP1 and GPC3, were first identified as key for cell proliferation in HCC. HCC generally spreads throughout the

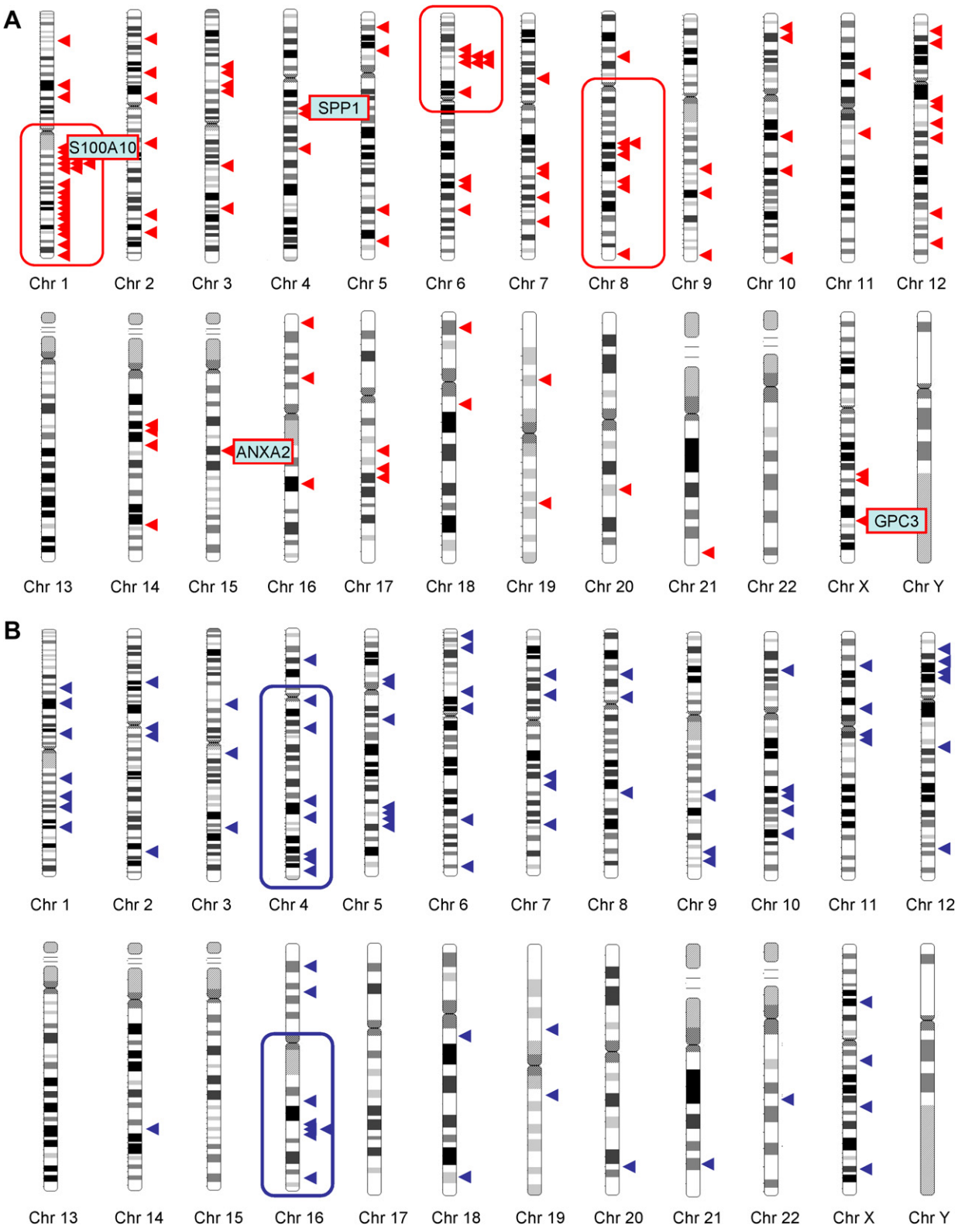

Fig. 5 – Location of 100 upregulated genes (A) and 100 downregulated genes (B) on human chromosomes. Upregulated genes are represented as red arrows, and downregulated genes are represented as blue arrows. Red-coloured circles (chromosomes 1q, 6p and 8q) represent regions with relative concentration of upregulated genes, whilst blue coloured circles (chromosomes 4q and 16q) represent regions with relative concentration of down-regulated genes. ANXA2, S100A10, SPP1 and GPC3 are located at chromosomes 15q21-22, 1q22, 4q21-25 and Xq26.1, respectively.

liver via the portal vein system even in advanced stages, and portal vein invasion is the most crucial histological feature associated with poor prognosis.[3] SPP1 protein expression is upregulated in primary HCC with accompanying metastasis, and SPP1 expression correlates with the invasiveness of HCC cells in tissue culture.[19] Based on network analysis, we speculated that the binding of integrins to SPP1[20] might be related to the progression and metastasis of HCC. GPC3, a heparan sulphate proteoglycan anchored to the plasma membrane, is also a good candidate marker of HCC. It is an oncofoetal protein overexpressed in HCC at both the mRNA and protein levels.[11] We also confirmed that GPC3 was overexpressed in HCC by the immunostaining of paraffin sections. In the activated integrin pathway, GPC3 interacts with IGF-2,[21] a protein that increases the phosphorylation of MAPK1.[22] GPC3 is also related to the zinc-finger transcription factors, GLI1 and GLI2, that are known players in WNT signalling and Sonic hedgehog signalling pathways.[23] Recently, WNT signalling was implicated in hepatocyte proliferation, which could be crucial in liver development, regeneration following partial hepatectomy, and pathogenesis of HCC.[24] In this context, SPP1 and GPC3 might participate in the activation of integrin signalling in HCC and based on the results of clustering analysis, might be implicated as mediators of intrahepatic metastasis, histopathological malignancy or poor prognosis.

Second, we focused on ANXA2, S100A10 and VIM, which were related to the Akt/NF-κB signalling pathway. ANXA2, also called calpactin I heavy chain, is a member of the annexin family of Ca^{2+}- and phospholipid-binding proteins and forms a heterotetrameric complex with S100A10, also called calpactin I light chain.[25] The ANXA2-S100A10 complex has been implicated in the structural organisation and dynamics of endosomal membranes, the organisation of cholesterol-rich membrane microdomains, and connecting lipid rafts with the actin cytoskeleton.[25] The ANXA2-S100A10 complex was also recently associated with recycling endosomes, and might be involved in the recycling of E-cadherin during the formation of the E-cadherin-based adherens junctions via the modulation of the actin cytoskeleton.[26] Moreover, ANXA2 was identified as a Rac binding partner and Rac activation is induced by the interactions of E-cadherin in the formation of adherens junctions.[27] In this way, cadherin–cadherin interactions initiate a cascade of signalling events that result in increased cadherin/Akt association, activation of Akt/NF-κB signalling, and increased cell survival and tumour growth.[27] Akt1 was found to associate structurally with VIM (a structural component of intermediate filaments),[28] which has been found in poorly differentiated HCC as well as hepatoblastomas.[29] Therefore, it is possible that binding of Akt1 and VIM activates downstream players (NF-κB signalling) as well as increasing the intrinsic activity of Akt1. This molecular understanding of HCC progression in Akt/NF-κB signalling was not so different from our result of correlations between clinicopathological features and gene expression profiles. It seems that the higher-activated group in Akt/NF-κB signalling has lower histopathological differentiation.

Currently, the array-based CGH approach is used to study chromosomal aberrations in human cancers. A previously reported meta-analysis[30] showed that the most common chromosomal arms containing gains were 1q, 6p and 8q, whereas the most common losses were found in chromosomes 4q, 8p and 16q. Comparing our expression data with the meta-analysis result of array-based CGH in HCC,[30] we found that our gene expression data surprisingly matched the chromosomal aberrations. The comprehensive analysis of 100 HCC samples using human 30 K DNA microarray revealed a potential association between the global copy number and expression. It is also noteworthy that our identified key molecules that operate synergistically in hepatocarcinogenesis are located at separate chromosomes, so chromosomal aberrations cannot prove a relationship of candidate genes such as ANXA2 and S100A10. Therefore, our integrative network approach can provide a significant clue to the discovery of novel genetic combinations that may be important for hepatocarcinogenesis.

Here, we highlighted the 'hotspot' canonical pathways in HCC and improved our molecular understanding of HCC progression. It is widely recognised that there are distinct molecular subtypes of HCC in the transcriptome space, and current interest of the community spread to include identification of subtype-specific aberration of genes/pathway. This functional genomics study could contribute towards the detection of several signalling pathways commonly activated in HCC. Moreover, we succeeded in detecting two potential disease markers, ANXA2 and S100A10, whose colocalisation in human HCC tissues has not been reported previously.

In conclusion, we reported an integrative approach of genome-wide microarray analysis and network analysis in HCC. This novel approach allows the extraction of deeper biological insight from microarray data and identifying potential key molecules in hepatocarcinogenesis.

Conflict of interest statement

None declared.

Acknowledgements

We thank EIJI MIYOSHI, Department of Biochemistry, Osaka University Medical School, and JORGE FILMUS, Division of Molecular and Cell Biology, Sunnybrook and Women's College Health Sciences Centre and Department of Medical Biophysics, University of Toronto, for providing a monoclonal antibody of GPC3.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejca.2008.02.019.

REFERENCES

1. Thomas MB, Abbruzzese JL. Opportunities for targeted therapies in hepatocellular carcinoma. *J Clin Oncol* 2005;**23**:8093–108.

2. Jemal A, Murray T, Ward E, Samuels A, Tiwari RC, Ghafoor A. Cancer statistics, 2005. *CA Cancer J Clin* 2005;**55**:10–30.
3. Pan HW, Ou YH, Peng SY, Liu SH, Lai PL, Lee PH. Overexpression of osteopontin is associated with intrahepatic metastasis, early recurrence, and poorer prognosis of surgically resected hepatocellular carcinoma. *Cancer* 2003;**98**:119–27.
4. Tsou AP, Wu KM, Tsen TY, Chi CW, Chiu JH, Lui WY. Parallel hybridization analysis of multiple protein kinase genes: identification of gene expression patterns characteristic of human hepatocellular carcinoma. *Genomics* 1998;**50**:331–40.
5. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V. Genetics of gene expression surveyed in maize, mouse and man. *Nature* 2003;**422**:297–302.
6. Brem RB, Yvert G, Clinton R, Kruglyak L. Genetic dissection of transcriptional regulation in budding yeast. *Science* 2002;**296**:752–5.
7. Thorgeirsson SS, Grisham JW. Molecular pathogenesis of human hepatocellular carcinoma. *Nat Genet* 2002;**31**:339–46.
8. Slaton JW, Perrotte P, Inoue K, Dinney CP, Fidler IJ. Interferon-alpha-mediated down-regulation of angiogenesis-related genes and therapy of bladder cancer are dependent on optimization of biological dose and schedule. *Clin Cancer Res* 1999;**5**:2726–34.
9. Yamamoto T, Nagano H, Sakon M, Wada H, Eguchi H, Kondo M. Partial contribution of tumour necrosis factor-related apoptosis-inducing ligand (TRAIL)/TRAIL receptor pathway to antitumor effects of interferon-alpha/5-fluorouracil against hepatocellular carcinoma. *Clin Cancer Res* 2004;**10**:7884–95.
10. Kondo M, Yamamoto H, Nagano H, Okami J, Ito Y, Shimizu J. Increased expression of COX-2 in nontumor liver tissue is associated with shorter disease-free survival in patients with hepatocellular carcinoma. *Clin Cancer Res* 1999;**5**:4005–12.
11. Capurro M, Wanless IR, Sherman M, Deboer G, Shi W, Miyoshi E. Glypican-3: a novel serum and histochemical marker for hepatocellular carcinoma. *Gastroenterology* 2003;**125**:89–97.
12. Zhang H, Ozaki I, Mizuta T, Yoshimura T, Matsuhashi S, Hisatomi A. Mechanism of beta 1-integrin-mediated hepatoma cell growth involves p27 and S-phase kinase-associated protein 2. *Hepatology* 2003;**38**:305–13.
13. Mottet D, Dumont V, Deccache Y, Demazy C, Ninane N, Raes M. Regulation of hypoxia-inducible factor-1alpha protein level during hypoxic conditions by the phosphatidylinositol 3-kinase/Akt/glycogen synthase kinase 3beta pathway in HepG2 cells. *J Biol Chem* 2003;**278**:31277–85.
14. Puisieux A, Ji J, Ozturk M. Annexin II up-regulates cellular levels of p11 protein by a post-translational mechanisms. *Biochem J* 1996;**313**:51–5.
15. Zobiack N, Gerke V, Rescher U. Complex formation and submembranous localization of annexin 2 and S100A10 in live HepG2 cells. *FEBS Lett* 2001;**500**:137–40.
16. Hollywood K, Brison DR, Goodacre R. Metabolomics: current technologies and future trends. *Proteomics* 2006;**6**:4716–23.
17. Cantley LC. The phosphoinositide 3-kinase pathway. *Science* 2002;**296**:1655–7.
18. Mootha VK, Bunkenborg J, Olsen JV, et al. Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* 2003;**115**:629–40.
19. Ye QH, Qin LX, Forgues M, He P, Kim JW, Peng AC. Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nat Med* 2003;**9**: 416–23.
20. Hu DD, Lin EC, Kovach NL, Hoyer JR, Smith JW. A biochemical characterization of the binding of osteopontin to integrins alpha v beta 1 and alpha v beta 5. *J Biol Chem* 1995;**270**: 26232–8.
21. Song HH, Shi W, Filmus J. OCI-5/rat glypican-3 binds to fibroblast growth factor-2 but not to insulin-like growth factor-2. *J Biol Chem* 1997;**272**:7574–7.
22. Moorehead RA, Sanchez OH, Baldwin RM, Khokha R. Transgenic overexpression of IGF-II induces spontaneous lung tumors: a model for human lung adenocarcinoma. *Oncogene* 2003;**22**:853–7.
23. Regl G, Kasper M, Schnidar H, Eichberger T, Neill GW, Ikram MS. The zinc-finger transcription factor GLI2 antagonizes contact inhibition and differentiation of human epidermal cells. *Oncogene* 2004;**23**:1263–74.
24. Apte U, Zeng G, Muller P, Tan X, Micsenyi A, Cieply B. Activation of Wnt/beta-catenin pathway during hepatocyte growth factor-induced hepatomegaly in mice. *Hepatology* 2006;**44**:992–1002.
25. Gerke V, Moss SE. Annexins: from structure to function. *Physiol Rev* 2002;**82**:331–71.
26. Yamada A, Irie K, Hirota T, Ooshio T, Fukuhara A, Takai Y. Involvement of the annexin II-S100A10 complex in the formation of E-cadherin-based adherens junctions in Madin–Darby canine kidney cells. *J Biol Chem* 2005;**280**: 6016–27.
27. Kovacs EM, Ali RG, McCormack AJ, Yap AS. E-cadherin homophilic ligation directly signals through Rac and phosphatidylinositol 3-kinase to regulate adhesive contacts. *J Biol Chem* 2002;**277**:6708–18.
28. Siu MK, Wong CH, Lee WM, Cheng CY. Sertoli-germ cell anchoring junction dynamics in the testis are regulated by an interplay of lipid and protein kinases. *J Biol Chem* 2005;**280**:25029–47.
29. Abenoza P, Manivel JC, Wick MR, Hagen K, Dehner LP. Hepatoblastoma: an immunohistochemical and ultrastructural study. *Hum Pathol* 1987;**18**: 1025–35.
30. Moinzadeh P, Breuhahn K, Stutzer H, Schirmacher P. Chromosome alterations in human hepatocellular carcinomas correlate with aetiology and histological grade–results of an explorative CGH meta-analysis. *Br J Cancer* 2005;**92**:935–41.