BIOINFORMATICS ORIGINAL PAPER

Vol. 29 no. 4 2013, pages 420–427
doi:10.1093/bioinformatics/bts715

Genome analysis

Advance Access publication December 24, 2012

In silico identification of oncogenic potential of fyn-related kinase in hepatocellular carcinoma

Jia-Shing Chen[1,†], Wei-Shiang Hung[1,†], Hsiang-Han Chan[1], Shaw-Jenq Tsai[2,3] and H. Sunny Sun[1,3,*]

[1]Institute of Molecular Medicine, College of Medicine, [2]Department of Physiology, College of Medicine and [3]Bioinformatics Center, Center for Biotechnology and Biosciences, National Cheng Kung University, Tainan 70101, Taiwan, Republic of China

Associate Editor: Alfonso Valencia

ABSTRACT

Motivation: Cancer development is a complex and heterogeneous process. It is estimated that 5–10% of human genes probably contribute to oncogenesis, whereas current experimentally validated cancer genes only cover 1% of the human genome. Thus hundreds of cancer genes may still remain to be identified. To search for new genes that play roles in carcinogenesis and facilitate cancer research, we developed a systematic workflow to use information saved in a previously established tumor-associated gene (TAG) database.

Results: By exploiting the information of conserved protein domains from the TAG, we identified 183 potential new TAGs. As a proof-of-concept, one predicted oncogene, fyn-related kinase (FRK), which shows an aberrant digital expression pattern in liver cancer cells, was selected for further investigation. Using 68 paired hepatocellular carcinoma samples, we found that FRK was up-regulated in 52% of cases ($P<0.001$). Tumorigenic assays performed in Hep3B and HepG2 cell lines revealed a significant correlation between the level of FRK expression and invasiveness, suggesting that FRK is a positive regulator of invasiveness in liver cancer cells.

Conclusion: These findings implied that FRK is a multitalented signal transduction molecule that produces diverse biological responses in different cell types in various microenvironments. In addition, our data demonstrated the accuracy of computational prediction and suggested that other predicted TAGs can be potential targets for future cancer research.

Availability: The TAG database is available online at the Bioinformatics Center website: http://www.binfo.ncku.edu.tw/TAG/.

Contact: hssun@mail.ncku.edu.tw

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on June 5, 2012; revised on November 21, 2012; accepted on December 18, 2012

1 INTRODUCTION

Hepatocellular carcinoma (HCC) is the fifth most common cancer worldwide (Caldwell and Park, 2009) and shows high prevalence in Asia and Africa (Bosch *et al.*, 2004). According to the annual report from the Bureau of Health Promotion, Department of Health, Taiwan (http://www.bhp.doh.gov.tw/bhpnet/portal/Default.aspx), HCC is the second most common cancer in the Taiwanese population but is the leading cause of cancer-related death in Taiwanese males. Although recent studies have revealed many genetic and epigenetic changes that lead to the aberrant activation of several signaling cascades, including epidermal growth factor receptor and Ras/extracellular signal-regulated kinase (Llovet and Bruix, 2008), the pathogenesis of HCC is heterogeneous among patients (Cha and Dematteo, 2005), so the molecular mechanisms of hepatocarcinogenesis remain largely unclear.

Cancer development is a complex and heterogeneous process and it may be brought about by multiple causes, including chemical and physical mutation, viral or bacterial infection and hereditary mutations (Carrillo-Infante *et al.*, 2007; Fasano and Muggia, 2009; Kuper *et al.*, 2002). Furthermore, cancer is considered to be a genetic disease and develops, at least in part, because of DNA damage (Vogelstein and Kinzler, 2004). Generally speaking, genes involved in tumorigenesis are divided into two basic categories based on their actions: oncogenes and tumor suppressor genes (TSGs) (Bowden *et al.*, 1994). While oncogenes encode proteins that control cell proliferation and/or apoptosis (Croce, 2008), a TSG is a protein that protects a cell from entering the course of carcinogenesis (Sherr, 2004). It has been suggested that 5–10% of human genes probably contribute to oncogenesis (Strausberg *et al.*, 2003), whereas current experimentally validated cancer genes only cover 1% of the human genome (Futreal *et al.*, 2004), suggesting that hundreds of cancer genes still remain to be identified.

A structural element is the signature of a protein that specifies the unique function of that protein. Typically, structural elements like domains or motifs provide clues to the functions of otherwise uncharacterized proteins and are often used to search for novel members of a protein family (Hawkins and Kihara, 2007). Many new proteins belonging to various families have been identified through *in silico* approaches in humans and other species (Masignani *et al.*, 2004; Schutte *et al.*, 2002). For example, based on the well-known five major cysteine-rich domains, Chen *et al.* (2004) found a novel gel-forming mucin (MUC19/Muc19) in glandular tissues using a hidden Markov model-based genome-wide search approach. Another example is the identification of new claudin family members with a PSI-BLAST-based approach (Wu *et al.*, 2006). These studies

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

demonstrate that computer-assisted approaches are efficient and flexible for identifying novel members of domain-sharing protein families.

We hypothesized that genes involved in tumorigenesis, especially functionally distinct groups of genes like oncogenes and TSGs, code for global protein families; thus the members within a group may share common domains that are signatures of their unique pathophysiological function. If so, the well-characterized tumor-associated protein domains can be used to search for novel proteins that may be involved in tumorigenesis. To test this hypothesis, we first established a tumor-associated gene (TAG) database (http://www.binfo.ncku.edu.tw/TAG/) to provide information about commonly shared functional domains in well-characterized oncogenes and TSGs (Chan *et al.*, 2007). Using TAG genes as a training set followed by computer-assisted searching in the human genome, we identified 183 candidates with the potential to be involved in tumorigenesis. For a proof-of-concept study, one candidate, Fyn-related kinase (FRK), was chosen and its oncogenic potential for invasiveness in HCC was validated.

2 MATERIAL AND METHODS

2.1 Architecture of tumor-associated gene database

All genes collected in the TAG database were identified through a text-mining approach. The comprehensive search used 'oncogene', 'tumor suppressor gene' and 'humans' as keywords and searched in the fields of 'Abstract' and 'Title' from 1970 to 2010 in the PubMed database (http://www.ncbi.nlm.nih.gov/pubmed/). A semi-automatic information retrieval engine was designed using Perl Script Language. The system collected specific information from various internet resources and stored it in the TAG database. In general, gene-related information was collected from Entrez Gene (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene), Gene Ontology (http://www.geneontology.org/) and GeneCards (http://www.genecards.org/), whereas protein-related information was collected from SwissProt (http://www.expasy.org/sprot/), InterPro (http://www.ebi.ac.uk/interpro/) and Blocks (http://blocks.fhcrc.org/).

The TAG was constructed using MySQL on the RedHat Linux (V7.3) operating system. The database consists of five tables: gene content, gene ontology, disease, protein domain and TAG domain weight-matrix. Each table summarizes attributes for a specific entity related to the TAGs. The original TAG database was established in 2007 (Chan *et al.*, 2007) and was updated in 2010 to include newly published genes involved in cancer development. Currently, the TAG database contains information on 246 oncogenes, 265 TSGs and 151 genes related to tumorigenesis.

2.2 Construction of weight-matrix table of TAG domains

To define the weight of a TAG domain, 246 oncogenes and 265 TSGs from the current TAG database were used as the training set in the following analysis. The 2D term frequency formula was adapted from the term frequency-inverse document frequency (tf-idf) and used to construct the weight-matrix table of TAG domains. The tf-idf weight is often used in information retrieval and text mining (Salton and Buckley, 1988). This weight is a statistical measure used to evaluate how important a word is to a document in a collection. The importance increases proportionally to the number of times a word appears in the document (i.e. *tf*) but is offset by the frequency of the word in the document (i.e. *df*). The original term frequency in a given document gives a measurement of the importance of the term within the particular document.

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

In this case, $n_{i,j}$ is the number of occurrences of a specified term i in the document d_j, and the denominator is the sum of the total number of occurrences of all terms in the document d_j. In our case, the 'term' represents each TAG domain and the 'document' is either an oncogene or TSG group. To evaluate whether a TAG domain contributing to the domain-containing protein functions as oncogene or TSG, we first counted the number of appearances of all TAG domains in the oncogene and TSG groups and listed them in a table. Then we used the frequencies of a given TAG domain within the group (i.e. oncogene) and across groups (i.e. oncogene and TSG) to represent the importance and specificity of each TAG domain, respectively. The mathematical 2D term frequency formula used to calculate the weight of a TAG domain was as follows:

$$W_{i,j} = \frac{n_{i,j}}{\sum_{k=1}^{\mathrm{Ij}} n_{k,j}} \times \frac{n_{i,j}}{\sum_{l=1}^{J} n_{i,l}}$$

where $W_{i,j}$ represents the weight for a domain in the ith row and jth column in the table, and $n_{i,j}$ is the number of occurrences of a specified domain i in the group j. While k is from 1 to I_j, and I_j stands for the number of different domains in the group j, l is from 1 to J, and J stands for the number of different groups. Under current settings in the weight-matrix table, there are only oncogene and TSG groups, so J equals 2. The denominator on the left is the sum of the total number of all domains in a group j. The denominator on the right is the sum of the total number of a specified domain i in all groups. The weight increases when the appearance of a given domain in the specified group is high and across multiple groups is low; therefore the weight tends to enhance the relative importance of a domain in one group *versus* the other. A weight-matrix table was generated based on the domain information from all TAGs, and each TAG domain has an oncogene weight and a TSG weight (Supplementary Table S1). The oncogenic potential of any protein is the score summed from all domains present in the given protein.

2.3 Computational identification of new TAGs

We analyzed the domain information, constructed a weight-matrix table and determined the cutoff values for defining 'oncogene' and 'TSG' using the training set data. The workflow for new TAG prediction (Fig. 1) is based on the human full-length cDNA (FLC) sequences retrieved from the NCBI GenBank database (http://www.ncbi.nlm.nih.gov/genbank/). All FLC sequences were then translated by the *transeq* tool (http://emboss.sourceforge.net/apps/cvs/transeq.html). The deduced amino-acid sequences were used to screen for the presence of TAG domains using InterProScan (http://www.ebi.ac.uk/InterProScan/). Following the procedure used to assign scores for proteins in the training sets, every individual amino-acid sequence had two scores, one each for the oncogene and TSG group. Individual sequences that had a score that passed the cutoff value for an oncogene or TSG were considered to be TAG candidates. The overall workflow for computational TAG identification is given in Figure 1.

To understand the properties and physiological significance of the predicted new TAGs, genes that were considered to be new TAGs were further analyzed for gene ontology using the MetaCore software suite (GeneGo, St Joseph, MI, USA). Significance probability was calculated using the hypergeometrical distribution based on gene ontology terms.

2.4 Specimens and cell lines

A total of 68 paired HCC samples (tumor and matched non-tumor) as well as four HCC cell lines (Hep3B, HepG2, PLC/PRF/5 and Huh7) were used to study the role of a candidate TAG in liver tumor development. All of the tissue samples were freshly frozen at −80°C until further analysis. The detailed information of both HCC patients and cell lines was given previously (Chen *et al.*, 2012).

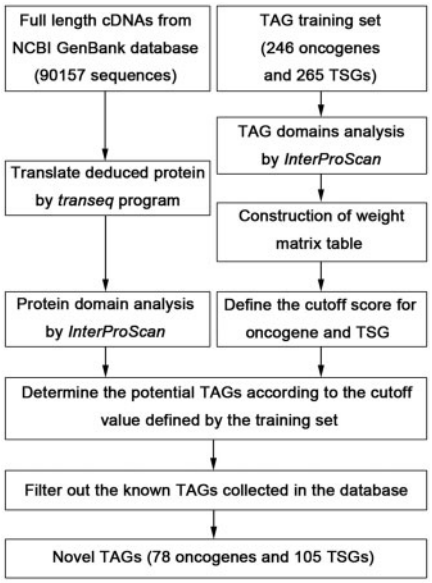

Fig. 1. Schematic of workflow for the identification of new TAGs

2.5 Detection of endogenous FRK

Multiple tissue cDNA panels (BioChain, Hayward, CA, USA) and liver cancer cell lines were used to detect the endogenous mRNA expression pattern of FRK using the primers FRK-F: 5′-GTCCCAGC TCCATTTGATTT-3′ and FRK-R: 5′-TTTTCACTGCTACTGGAG TGGT-3′. In addition, total protein prepared from cultured cells and paired HCC samples was subjected to western blot analysis.

2.6 Transient overexpression of FRK and RNA interference

FRK-overexpressing and pcDNA3.1+ plasmids were from Origene (Rockville, MD, USA) and Invitrogen (Carlsbad, CA, USA), respectively. The lentiviral short hairpin RNA (shRNA) clones that targeted the Luc or FRK genes were from the National RNAi Core Facility of Academia Sinica, Taiwan. The experiments on overexpression and knockdown of FRK in HepG2 and Hep3B cells were performed with a procedure similar to that described previously (Chen *et al.*, 2012). Overexpression and knockdown of FRK protein were confirmed by western blotting. All knockdown and transient overexpression experiments were performed in triplicate and at least three times independently.

2.7 Cell proliferation assay

The Hep3B and HepG2 cells were transfected with FRK construct and pcDNA3.1+, respectively, in 60-mm dishes. Six hours after transfection, trypsinized cells were suspended in culture medium and $\sim 5 \times 10^3$ cells were seeded in 96-well plates (each group had four wells). Cell proliferation was assessed using a colorimetric assay (CellTiter 96® AQueous One Solution cell proliferation assay; Promega, Madison, WI, USA) at various time points. The color intensity was measured at 490 nm using a 96-well plate reader (Labsystems, Multiskan EX, Helsinki, Finland). The corrected absorbance (after subtracting control blanks) was used to determine the proliferative response. All experiments were repeated independently at least three times.

2.8 Cell invasion assay

The invasion assay was performed using transwells with an 8-μm pore size polycarbonate membrane filter (Millipore, Temecula, CA, USA) coated with a uniform layer of 15.15 μg/cm^2 Matrigel basement membrane matrix. Before carrying out this experiment, the Matrigel was reconstituted by adding 40 μl culture medium with 0.1% fetal bovine serum (FBS) and incubated at 37°C for 1 h. Roughly 1.5×10^5 trypsinized cells were suspended in 100 μl culture medium with 0.1% FBS and added into the upper chamber of the transwell. Culture medium with 10% FBS (700 μl) was added to the lower chamber. After incubation for 20 h, the non-invading cells and the upper side of the membrane were removed with a cotton swab. The cells on the lower surface of the membrane were fixed in methanol for 15 min and stained with 0.2% crystal violet for 30 min. The invading cells were counted by dividing the membrane area into five fields and visualizing the fields at 200× magnification. The examinations were in triplicates and the experiments were repeated at least three times independently.

2.9 Colony formation assay

In the colony formation assay, cells were suspended in 0.3% agar with complete culture medium and plated at a density of 2×10^4 cells per 60-mm dish pre-coated with 0.5% agar. Medium (0.5 ml) was added every 3 days and incubated at 37°C. After 4 weeks, the colonies were stained with 0.2% crystal violet (Fisher Scientific, Pittsburgh, PA, USA) for >1 h. Colonies >100 μm in diameter were scored under a microscope at 40× magnification. The examinations were in triplicate plates and a total of 16 fields were counted for each plate. The experiments were repeated at least three times.

2.10 Statistical analysis

All data are presented as mean ± standard error. The means of groups were compared with the paired *t*-test using GraphPad Prism 5 (GraphPad Software, Inc. CA, USA). The receiver–operator characteristic (ROC) curve was analyzed using SPSS for Windows version 17.0. For all analyses, $P < 0.05$ was considered statistically significant. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

3 RESULTS

3.1 Defining cutoff scores for oncogenes and TSGs

The TAG database contained 366 domains (Supplementary Table S1). Among these, 49 were present in both the oncogene and TSG groups, thus having a weight for each group. The remaining 317 domains were uniquely in either the oncogene or the TSG group. Each TAG gene in the training set had an oncogene and a TSG score; these were calculated as the sum of the oncogene weights and TSG weights of the domains in that gene. After multiplying by 100, the oncogene and TSG scores of genes in the TAG database ranged from 0.1 to 18.08 (average, 3.71) and from 0.05 to 3.21 (average, 0.86), respectively.

The accuracy of the prediction model was assessed by plotting the ROC curve and measuring the area under the curve (AUROC). The AUROC for 'oncogene' was 0.85 and for 'TSG' was 0.80, indicating that both were excellent predictive models (Supplementary Fig. S1). To define the cutoff value for further analysis, we used the cutoff points from ROC analysis to minimize the false-positive rate (i.e. maximize specificity; 100%). The score for the 'oncogene' group signature was 9.09 and for 'TSG' it was 2.31. We found that 34 oncogenes and 12 TSGs in the original training set passed these cutoff values, representing 17 and 7% sensitivity in prediction using training set data. These results indicate that the defined cutoff values are stringent and should be able to predict genes with unique domain signatures.

3.2 Identification of new TAGs by computational prediction

A total of 90 157 FLC sequences were downloaded and analyzed by the procedure illustrated in Figure 1. We found 33 959 transcripts containing TAG domains with domain numbers ranging from 1 to 11 (Supplementary Table S2). This included 8958 unique gene symbols and 498 transcripts currently without symbols. The results of TAG domain analysis of FLC were filtered by the TAG training set to identify novel TAGs. Using the cutoff values for the 'unique domain signature' determined by the training set, 354 and 269 transcripts representing 119 and 125 unique gene symbols were predicted to be TAGs in the oncogene and TSG groups, respectively (Supplementary Tables S3 and S4). After removing the current TAGs from the predicted gene list, we finally obtained 183 new TAGs that included 78 oncogenes and 105 TSGs.

To reveal the molecular properties of these TAGs, we applied pathway and gene ontology analysis using the MetaCore software package. Among the 78 new TAG oncogenes, most were functionally involved in protein kinase activity (71 objects, *P*-value 1.047E-105, Supplementary Fig. S2A) and ~60% of these proteins were specifically implicated in protein tyrosine kinase (PTK) activity (*P*-value 2.444E-82). Thus, these predicted oncogenes are involved in many tumorigenic pathways including calcium signaling and mitogenic signaling (Fig. 2A). Furthermore, 19 out of 26 significantly associated diseases (*P*-value <1.0E-10) were either tumors/carcinomas or neoplasms, indicating that dysfunction of these genes indeed was associated with cancer formation (71 objects, *P*-value 1.047E-105, Supplementary Fig. S2B) and clustered on ovarian, breast and lung cancers (Disease networks, Fig. 2B). On the other hand, the properties of predicted TSGs were diverse, except that many exhibited binding ability (84 objects, *P*-value 9.323E-07, Supplementary Fig. S2C) and were involved in responses to stimuli (73 objects, *P*-value 3.257E-10). The most significant process according to the GeneGo Map categories for TSG was vascular development-angiogenesis (Fig. 2C), and dysfunction of these predicted new TSGs is known to contribute to many diseases like arteriosclerosis (21 objects, *P*-value 3.532E-09, Supplementary Fig. S2D), and may lead to the development of cancers from various tissues including prostate, lung and colon (Fig. 2D). In summary, these data strongly support roles of the predicted TAGs in tumorigenesis.

As proof-of-principle, the next step was to seek further experimental validation of the predicted TAGs. The likely functions and expression patterns of predicted TAGs were evaluated using information retrieved from the OMIM, GeneCard and PubMed databases. Because liver cancer is one of the most common malignancies in Taiwan and we have been working on identifying new biomarkers for HCC (Chen *et al.*, 2012), we selected one predicted oncogene, FRK, which functions as a non-receptor tyrosine kinase and shows an aberrant digital expression pattern in liver cancer cells (GeneCard; http://www.genecards.org/), for further investigation.

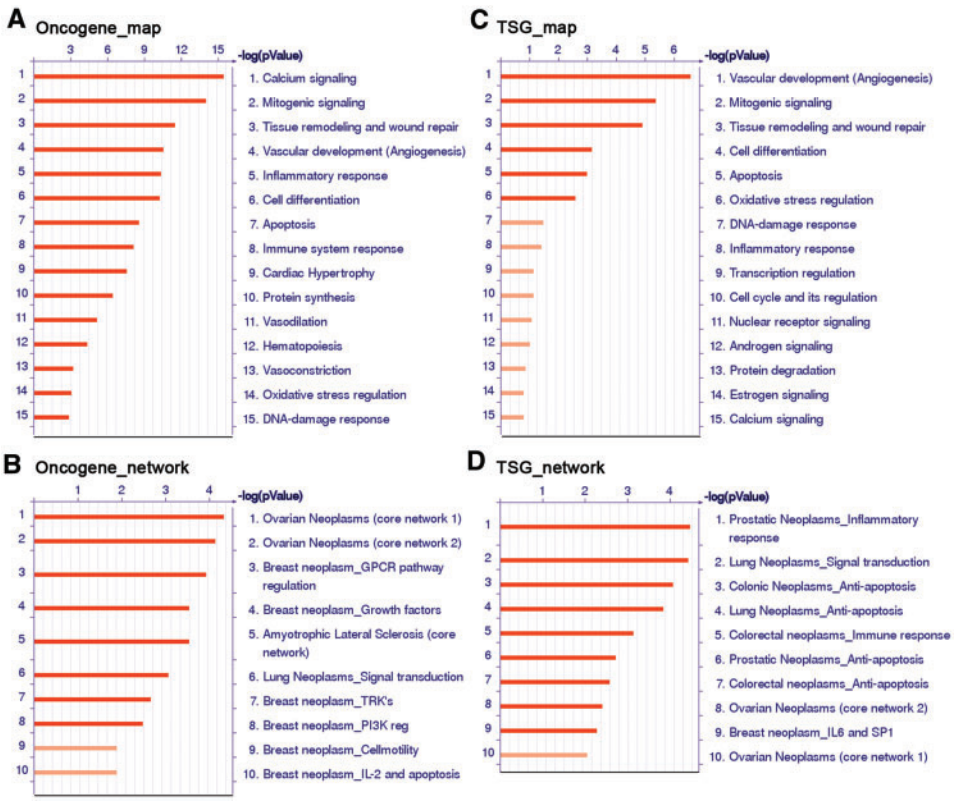

Fig. 2. Histograms representing the top GeneGo Map Folders and Disease Networks from analysis of the gene lists of newly predicted oncogenes (**A** and **B**) and tumor suppressors (**C** and **D**). Key differentially representative processes are shown. In each histogram, the longer each bar the more significant (−log pValue), and both map folders (A and C) and disease networks (B and D) are organized according to descending significance

3.3 Increased expression of endogenous FRK in liver cancer cells

To determine whether FRK dysfunction is involved in liver cancer progression, we first assessed FRK expression using commercially available cDNA panels prepared from both normal and tumor cells of liver, lung and colon. The results from RT-PCR showed that FRK was expressed in liver tumor and normal lung tissues (Fig. 3A). Next, we assessed FRK expression in four HCC cell lines. Using RT-PCR and western blot analyses, we found that both FRK mRNA and protein (Fig. 3B) were highly expressed in these lines. Further assessment of FRK expression in liver cancer cells was carried out in 68 paired HCC samples. Western blotting showed an average 1.6-fold increase in FRK expression (ranging from 1.03- to 4.14-fold) in 52% (35/68) of the tumor samples ($P<0.001$, Fig. 3C and Supplementary Table S5). Taken together, endogenous FRK expression showed up-regulation in HCC cells and this implied a functional role of FRK in hepatocarcinogenesis.

3.4 Higher levels of FRK expression in HCC patients without viral infection

The correlations between FRK-expression levels and various clinical features were analyzed. The levels were similar in males and females ($P=0.195$; Supplementary Fig. S3A) and were not associated with cirrhosis status ($P=0.341$; Supplementary Fig. S3B). Furthermore, FRK expression was independent of tumor size ($P=0.469$; Supplementary Fig. S3C), tumor type ($P=0.218$; Supplementary Fig. S3D) and stage of clinical pathology ($P=0.1$; Supplementary Fig. S3E). Nevertheless, a significant association between FRK expression and status of viral infection was found. Compared with patients without HBV or HCV infection, patients with HBV infection tended to express lower amounts of FRK ($P<0.05$; Supplementary Fig. S3F). Although the difference did not reach statistical significance, a similar trend was observed for FRK expression in patients with HCV infection ($P=0.063$; Supplementary Fig. S3F). Collectively, these data suggested that endogenous FRK expression plays a predominant role in non-virus-infected HCC.

3.5 FRK expression promotes Hep3B but not HepG2 cell growth and transformation

To investigate the role of FRK in the carcinogenesis of HCC, several tumorigenic assays were carried out in FRK-overexpressing and -knockdown liver cancer cells. In transient overexpression experiments, the transfection efficiency of the recombinant FRK construct was 50–60% in Hep3B and 20–30% in HepG2 cells (Supplementary Fig. S4), and the expression of recombinant FRK was confirmed by western blotting in both cell lines (Fig. 4A). In addition, endogenous expression of FRK was knocked down by infection with lentivirus containing FRK-specific shRNA. Western blot analysis showed that the level of FRK expression dropped to 67% in Hep3B and 80% in HepG2 cells (Fig. 4B).

We used a cell proliferation assay to study the effect of FRK on growth in FRK-overexpressing and -knockdown cells. Results from the MTS assay revealed that FRK overexpression promoted the proliferation of Hep3B cells (Fig. 4C; $P<0.05$).

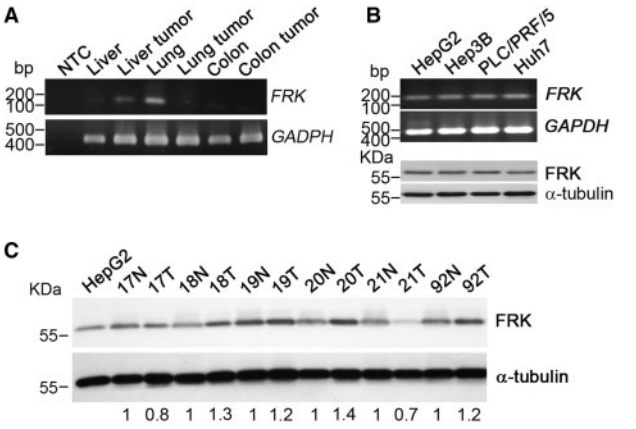

Fig. 3. Expression of FRK in HCC cell lines and tissues. (**A**) RT-PCR analysis of FRK mRNA expression in liver, liver tumor, lung, lung tumor, colon and colon tumor. GAPDH was used as internal control. (**B**) Expression of FRK mRNA (upper) by RT-PCR analysis and protein (lower) by western blot in four HCC cell lines. GAPDH and α-tubulin were used as loading controls. (**C**) Western blot analysis showing FRK protein expression in six paired HCC cases. HepG2 and α-tubulin were used as positive and loading controls. Numbers at the bottom of each lane represent relative expression level in tumor compared with normal tissue

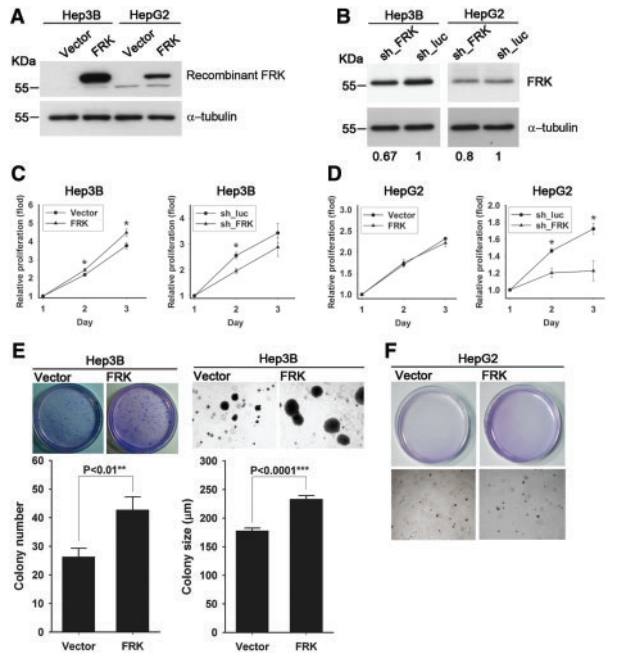

Fig. 4. FRK promotes proliferation in Hep3B cells but not in HepG2 cells. (**A**) Expression of recombinant FRK determined in Hep3B and HepG2 cells by western blot. (**B**) Knockdown of endogenous FRK analyzed in Hep3B and HepG2 cells by western blot. In both (A) and (B), α-tubulin was used as loading control. (**C**) Relative proliferation rates of Hep3B cells plotted for those carrying control vector or overexpressing FRK (left), and those containing control sh_luc or sh_FRK (right). (**D**) Relative proliferation rates of HepG2 cells carrying control vector or overexpressing FRK (left), and cells containing control sh_luc or sh_FRK (right). (**E**) Number (left) and size (right) of colonies in transformed Hep3B cells. Both transformed plates (upper) and barchart-s(lower) are shown. (**F**) Number (left) and size (right) of colonies in transformed HepG2 cells. Transformed plates (upper) and photographs of colonies (lower) are shown. $*P<0.05$, $**P<0.01$, $***P<0.001$

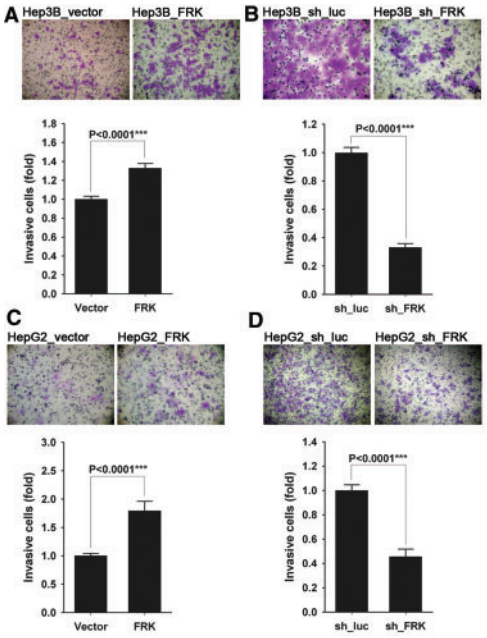

Fig. 5. FRK promotes invasion in Hep3B and HepG2 cell lines. The invasion assays were carried out using Matrigel-coated transwells. The crystal violet-stained cells (upper) were counted and the total number of cells from five randomly selected fields on the transwell were measured; the ratio of invasive cells compared with control cells is shown (lower). (**A**) Hep3B_vector and Hep3B_FRK cells, (**B**) Hep3B_sh_luc and Hep3B_sh_FRK cells, (**C**) HepG2_vector and HepG2_FRK cells and (**D**) HepG2_sh_luc and HepG2_sh_FRK cells. ***$P<0.001$

Furthermore, although the difference in growth rate did not reach the cutoff for significance, FRK knockdown tended to reduce proliferation in Hep3B compared with sh_luc control cells (Fig. 4C). Interestingly, FRK overexpression had no effect on proliferation but showed a noteworthy growth reduction ($P<0.05$) in FRK-knockdown HepG2 cells (Fig. 4D). Furthermore, transformation was tested in FRK-overexpressing Hep3B and HepG2 cells with the soft agar assay, which showed that enforced FRK expression promoted colony formation ($P<0.01$) and increased the colony size ($P<0.0001$) in Hep3B_FRK cells compared with the control cells (Fig. 4E). In contrast, no differences were found between FRK-overexpressing and vector-control HepG2 cells in either the number or size of colonies formed (Fig. 4F). Results from the transformation assay were in agreement with those of the proliferation assay and further confirmed that overexpression of FRK promoted cell growth in Hep3B but not HepG2 cells. All together, the findings indicated that FRK has oncogenic potential to regulate cell proliferation and transformation but the effect may vary in different cellular backgrounds or environments.

3.6 FRK expression enhances invasive ability in Hep3B and HepG2 cells

To determine whether FRK is involved in cell invasion, Matrigel-coated transwells were used to evaluate the invasiveness of FRK-overexpressing and FRK-knockdown cells. The results demonstrated that overexpressing FRK in Hep3B cells increased the number of invading cells ($P<0.0001$, Fig. 5A). In contrast, FRK knockdown strongly repressed the invasiveness of Hep3B cells ($P<0.0001$, Fig. 5B). Furthermore, HepG2_FRK showed increased and HepG2_sh_FRK cells showed reduced numbers of invasive cells compared with controls ($P<0.0001$ for both; Fig. 5C and D). Taken together, these data showed that altered expression of FRK correlated well with invasiveness in both HCC cell lines, thus strongly supporting the involvement of FRK in the regulation of cell invasion.

4 DISCUSSION

Protein kinase is an enzyme that modifies other proteins by phosphorylation of tyrosine or serine/threonine amino-acid residues. This usually results in a functional change of the target protein. PTKs, which can be classified into receptor and non-receptor types, are a large and diverse multigene family involved in the regulation of a range of cellular processes, including proliferation, migration, differentiation, death and cell-to-cell communication in multicellular organisms (Hunter, 2009). In humans, perturbation of PTK signaling by mutations and other genetic alterations results in deregulated kinase activity and leads to the development of many diseases (Robertson *et al.*, 2000) and malignant transformation (Blume-Jensen and Hunter, 2001). Although the products of oncogenes can be classified into six broad groups (transcription factors, chromatin remodelers, growth factors, growth factor receptors, signal transducers and apoptosis regulators; Croce, 2008), >70% of the known oncogenes and proto-oncogenes involved in cancers code for PTKs (Yim *et al.*, 2009b). Thus it was not surprising that the computational search using known oncogenes as the training set showed that >50% of the predicted TAG oncogenes belonged to this specific family. And the consequences of dysfunction of these PTKs no doubt would be tumor formation. On the other hand, a tumor suppressor is a protein that either has a dampening or repressive effect on the regulation of the cell cycle or promotes apoptosis, and sometimes both, so it protects a cell from entering the path to carcinogenesis (Sherr, 2004). The functions of tumor-suppressor proteins fall into many categories including proteins that are responsible for DNA damage (Markowitz, 2000) and involved in cell adhesion (Hirohashi and Kanai, 2003). In agreement with these principles, we found that the properties of the predicted TSGs were mainly binding ability and involvement in the processes of responses to stimuli. As the most significant process involving these predicted TSGs was vascular development-angiogenesis, and angiogenesis is considered to be an important step toward metastasis, these data suggested that the newly predicted TSGs may participate in the control of tumor metastasis.

FRK (also known as RAK) is a 54-kDa PTK that belongs to the Src non-receptor tyrosine kinase family (Lee *et al.*, 1994) and shares about 50% amino-acid identity with Fyn, Src and Lyn (Cance *et al.*, 1994; Lee *et al.*, 1994; Meyer *et al.*, 2003). Because FRK has different structural features that comprise a putative bipartite nuclear localization signal in the SH2 domain (Cance *et al.*, 1994; Chandrasekharan *et al.*, 2002) and is uniquely localized in the nucleus (Cance *et al.*, 1994; Pendergast, 1996; Yim *et al.*, 2009a), FRK is known to have a different set of substrates from other Src family members. FRK is expressed

predominantly in epithelial tissues such as kidney, liver and lung (Cance *et al.*, 1994). Previous studies showed that overexpression of FRK suppresses growth through an association with pRb during the G1 and S phases (Craven *et al.*, 1995; Meyer *et al.*, 2003). Another study demonstrated that ectopic expression of FRK in breast cancer cell lines suppresses cell growth, invasion and colony formation *in vitro* and tumor growth *in vivo* (Yim *et al.*, 2009b). These findings demonstrated that FRK expression stabilizes PTEN protein and activates FRK-pRb–induced growth inhibition (Craven *et al.*, 1995; Yim *et al.*, 2009a). Nevertheless, one study has reported that the *t*(6;12)(q21;p13) translocation fuses ETV6 and FRK together and the fusion protein contributed directly to leukemogenesis in a patient with acute myelogenous leukemia (Hosoya *et al.*, 2005). In addition, insertionally mutagenized and overexpressed frk is frequently found in myeloblastosis-associated virus (MAV)-2-induced lung sarcomas in chickens (Pajer *et al.*, 2009). Furthermore, miR-330, a known regulator of FRK expression, is down-regulated in HCC and has been proposed to participate in hepatocarcinogenesis (Varnholt *et al.*, 2008). These studies along with our finding that FRK was overexpressed in 52% of HCC samples, strongly suggested the oncogenic potential of FRK. All together, these findings indicated that FRK is involved in tumor progression but the effect may be tissue- or cell-type specific.

Our clinical investigation demonstrated that FRK is involved in liver cancer carcinogenesis. Follow-up tumorigenic assays showed that FRK promoted growth and transformation in Hep3B but not in HepG2 cells. Both Hep3B and HepG2 are commonly used HCC cell lines, however, previous studies demonstrated that they have different properties. For example, Hep3B cells have lost two TSGs, the Rb and p53 genes, but HepG2 retains intact expression of these genes; thus Hep3B is more malignant than HepG2 (Puisieux *et al.*, 1993). As earlier studies have demonstrated that the FRK suppression of cell proliferation is associated with the presence of Rb (Craven *et al.*, 1995; Meyer *et al.*, 2003), this explains the discrepancy of cell proliferation/transformation assays between Hep3B and HepG2 with FRK overexpression. Because studies showed that FRK inhibits PTEN destabilization and inactivates pRb-induced growth inhibition, these data suggest that FRK is a negative regulator of cell proliferation, at least in breast and liver cancer cells.

Interestingly, we found that the expression levels of FRK were significantly correlated with invasiveness in both Hep3B and HepG2 cells despite their different genetic backgrounds. Previous studies showed that FRK activates the Ras-related protein 1 (RAP1) pathway and induces neurite outgrowth in PC12 cells, probably through a mechanism involving physical association with the Src homology 2 domain-containing adaptor protein B (SHB), increased SHB phosphorylation and FAK expression and increased binding of the CRKII-C3G complex (Anneren *et al.*, 2000, 2003). RAP1 was shown to promote vascular endothelial growth factor receptor 2 (VEGFR2) activation and angiogenesis *via* integrin $\alpha(v)\beta3$ in endothelial cells (Lakshmikanthan *et al.*, 2011), and activation of RAP1 increases prostate cancer cell migration and metastasis (Bailey *et al.*, 2009). We propose that FRK–SHB–RAP1 pathway may be involved in the increased invasiveness of HCC cells. Alternatively, a recent study reported that FRK reduces migration and invasion in human glioma cells by regulating JNK/c-Jun signaling (Zhou *et al.*, 2012). As the standard tumorigenic assays performed in this study were just for proof-of-principle, these data were not sufficient to demonstrate the mechanism underlying the FRK-mediated oncogenic effect. Further experiments as well as animal models will be required to fully illustrate the function of FRK in cancers. Yet, these results indicate that FRK is a versatile signal transduction molecule that induces diverse biological responses in different cell types under various conditions and tissue-specific microenvironments.

In conclusion, although there are a few databases with collections showing information on cancer-related genes, most of them are restricted to certain types of cancers or genes (e.g. the breast cancer database, http://www.itb.cnr.it/breastcancer/index.html or the IARC TP53 database, http://www-p53.iarc.fr/) and none provide predictive function in the way the TAG database does. This feature of TAG distinguishes it from other current online/commercially available bioinformatics tools. As the oncogenic potential of FRK was suggested by computational analysis, confirmation of the FRK effect on liver cancer provided strong support for the idea that predicted TAGs are highly likely to be involved in tumorigenesis. Finally, our analysis revealed that 183 candidates had the potential to trigger tumor development at various stages and/or in various tissues, and we believe the outcome of this study will benefit cancer research. Although this list of genes contributes significantly to our current knowledge of TAGs, the roles and specificity of these predicted TAGs in tumorigenesis require further experimental validation.

ACKNOWLEDGEMENTS

The authors thank the National Core Facilities of the NRPGM-Tumor Tissue Bank in Southern Taiwan and the Taiwan Liver Cancer Network for providing the HCC samples, and the National RNAi Core Facility, Taiwan, for various RNAi constructs. In addition, we would like to express our appreciation to Dr Iain C Bruce for reading the manuscript and Dr Shuen-Lin Jeng for statistical consultation.

Funding: This research is supported by grant [NSC 95-3112-B-006 -006] (to H.S.S.) from the National Science Council, Taiwan.

Conflict of Interest: none declared.

REFERENCES

Anneren,C. *et al.* (2000) GTK, a Src-related tyrosine kinase, induces nerve growth factor-independent neurite outgrowth in PC12 cells through activation of the Rap1 pathway. Relationship to Shb tyrosine phosphorylation and elevated levels of focal adhesion kinase. *J. Biol. Chem.*, **275**, 29153–29161.

Anneren,C. *et al.* (2003) The FRK/RAK-SHB signaling cascade: a versatile signal-transduction pathway that regulates cell survival, differentiation and proliferation. *Curr. Mol. Med.*, **3**, 313–324.

Bailey,C.L. *et al.* (2009) Activation of Rap1 promotes prostate cancer metastasis. *Cancer Res.*, **69**, 4962–4968.

Blume-Jensen,P. and Hunter,T. (2001) Oncogenic kinase signalling. *Nature*, **411**, 355–365.

Bosch,F.X. *et al.* (2004) Primary liver cancer: worldwide incidence and trends. *Gastroenterology*, **127**, S5–S16.

Bowden,G.T. *et al.* (1994) Oncogene activation and tumor suppressor gene inactivation during multistage mouse skin carcinogenesis. *Cancer Res.*, **54**, 1882s–1885s.

Caldwell,S. and Park,S.H. (2009) The epidemiology of hepatocellular cancer: from the perspectives of public health problem to tumor biology. *J. Gastroenterol.*, **44** (**Suppl. 19**), 96–101.
Cance,W.G. *et al.* (1994) Rak, a novel nuclear tyrosine kinase expressed in epithelial cells. *Cell Growth Differ.*, **5**, 1347–1355.
Carrillo-Infante,C. *et al.* (2007) Viral infections as a cause of cancer (review). *Int. J. Oncol.*, **30**, 1521–1528.
Cha,C. and Dematteo,R.P. (2005) Molecular mechanisms in hepatocellular carcinoma development. *Best Pract. Res. Clin. Gastroenterol.*, **19**, 25–37.
Chan,H.H. *et al.* (2007) TAG: a comprehensive database for cancer research. *NAR.,* **35**, Molecular Biology Database Collection, #992.
Chandrasekharan,S. *et al.* (2002) Characterization of mice deficient in the Src family nonreceptor tyrosine kinase Frk/rak. *Mol. Cell Biol.*, **22**, 5235–5247.
Chen,J.S. *et al.* (2012) Expression of T-cell lymphoma invasion and metastasis 2 (TIAM2) promotes proliferation and invasion of liver cancer. *Int. J. Cancer*, **130**, 1302–1313.
Chen,Y. *et al.* (2004) Genome-wide search and identification of a novel gel-forming mucin MUC19/Muc19 in glandular tissues. *Am. J. Respir. Cell Mol. Biol.*, **30**, 155–165.
Craven,R.J. *et al.* (1995) The nuclear tyrosine kinase Rak associates with the retinoblastoma protein pRb. *Cancer Res.*, **55**, 3969–3972.
Croce,C.M. (2008) Oncogenes and cancer. *N. Engl. J. Med.*, **358**, 502–511.
Fasano,J. and Muggia,F. (2009) Breast cancer arising in a BRCA-mutated background: therapeutic implications from an animal model and drug development. *Ann. Oncol.*, **20**, 609–614.
Futreal,P.A. *et al.* (2004) A census of human cancer genes. *Nat. Rev. Cancer*, **4**, 177–183.
Hawkins,T. and Kihara,D. (2007) Function prediction of uncharacterized proteins. *J. Bioinform. Comput. Biol.*, **5**, 1–30.
Hirohashi,S. and Kanai,Y. (2003) Cell adhesion system and human cancer morphogenesis. *Cancer Sci.*, **94**, 575–581.
Hosoya,N. *et al.* (2005) Identification of a SRC-like tyrosine kinase gene, FRK, fused with ETV6 in a patient with acute myelogenous leukemia carrying a t(6;12)(q21;p13) translocation. *Genes Chromosomes Cancer*, **42**, 269–279.
Hunter,T. (2009) Tyrosine phosphorylation: thirty years and counting. *Curr. Opin. Cell Biol.*, **21**, 140–146.
Kuper,H. *et al.* (2002) Tobacco use and cancer causation: association by tumour type. *J. Intern. Med.*, **252**, 206–224.
Lakshmikanthan,S. *et al.* (2011) Rap1 promotes VEGFR2 activation and angiogenesis by a mechanism involving integrin alphavbeta(3). *Blood*, **118**, 2015–2026.
Lee,J. *et al.* (1994) Cloning of FRK, a novel human intracellular SRC-like tyrosine kinase-encoding gene. *Gene*, **138**, 247–251.
Llovet,J.M. and Bruix,J. (2008) Molecular targeted therapies in hepatocellular carcinoma. *Hepatology*, **48**, 1312–1327.
Markowitz,S. (2000) DNA repair defects inactivate tumor suppressor genes and induce hereditary and sporadic colon cancers. *J. Clin. Oncol.*, **18**, 75S–80S.
Masignani,V. *et al.* (2004) In silico identification of novel bacterial ADP-ribosyltransferases. *Int. J. Med. Microbiol.*, **293**, 471–478.
Meyer,T. *et al.* (2003) Breast cancer cell line proliferation blocked by the Src-related Rak tyrosine kinase. *Int. J. Cancer*, **104**, 139–146.
Pajer,P. *et al.* (2009) Industasis, a promotion of tumor formation by nontumorigenic stray cells. *Cancer Res.*, **69**, 4605–4612.
Pendergast,A.M. (1996) Nuclear tyrosine kinases: from Abl to WEE1. *Curr. Opin. Cell Biol.*, **8**, 174–181.
Puisieux,A. *et al.* (1993) Retinoblastoma and p53 tumor suppressor genes in human hepatoma cell lines. *FASEB J.*, **7**, 1407–1413.
Robertson,S.C. *et al.* (2000) RTK mutations and human syndromes: when good receptors turn bad. *Trends Genet.*, **16**, 368.
Salton,G. and Buckley,C. (1988) Term-weighting approaches in automatic text retrieval. *Inf. Process. Manag.*, **24**, 513–523.
Schutte,B.C. *et al.* (2002) Discovery of five conserved beta -defensin gene clusters using a computational search strategy. *Proc. Natl Acad. Sci. USA*, **99**, 2129–2133.
Sherr,C.J. (2004) Principles of tumor suppression. *Cell*, **116**, 235–246.
Strausberg,R.L. *et al.* (2003) Sequence-based cancer genomics: progress, lessons and opportunities. *Nat. Rev. Genet.*, **4**, 409–418.
Varnholt,H. *et al.* (2008) MicroRNA gene expression profile of hepatitis C virus-associated hepatocellular carcinoma. *Hepatology*, **47**, 1223–1232.
Vogelstein,B. and Kinzler,K.W. (2004) Cancer genes and the pathways they control. *Nat. Med.*, **10**, 789–799.
Wu,J. *et al.* (2006) Identification of new claudin family members by a novel PSI-BLAST based approach with enhanced specificity. *Proteins*, **65**, 808–815.
Yim,E.K. *et al.* (2009a) Rak functions as a tumor suppressor by regulating PTEN protein stability and function. *Cancer Cell*, **15**, 304–314.
Yim,E.K. *et al.* (2009b) Exploring Rak tyrosine kinase function in breast cancer. *Cell Cycle*, **8**, 2360–2364.
Zhou,X. *et al.* (2012) FRK controls migration and invasion of human glioma cells by regulating JNK/c-Jun signaling. *J. Neurooncol.*, **2012**, 13.