

PROCEEDINGS

Open Access

Application of microRNA and mRNA expression profiling on prognostic biomarker discovery for hepatocellular carcinoma

Lin Wei^{1,2}, Baofeng Lian^{3,4}, Yuannv Zhang², Wei Li³, Jianren Gu², Xianghuo He^{2*}, Lu Xie^{3*}

From The Twelfth Asia Pacific Bioinformatics Conference (APBC 2014)
Shanghai, China. 17-19 January 2014

Abstract

Background: Hepatocellular carcinoma (HCC) is one of the most highly malignant and lethal cancers of the world. Its pathogenesis has been reported to be multi-factorial, and the molecular carcinogenesis of HCC can not be attributed to just a few individual genes. Based on the microRNA and mRNA expression profiling of normal liver tissues, pericancerous hepatocellular tissues and hepatocellular carcinoma tissues, we attempted to find prognosis related gene sets for HCC patients.

Results: We identified differentially expressed genes (DEG) from three comparisons: Cancer/Normal, Cancer/Pericancerous and Pericancerous/Normal. GSEA (gene set enrichment analysis) were performed. Based on the enriched gene sets of GO terms, pathways and transcription factor targets, it was found that the genome instability and cell proliferation increased while the metabolism and differentiation decreased in HCC tissues. The expression profile of DEGs in each enriched gene set was used to correlate to the postoperative survival time of HCC patients. Nine gene sets were found to prognostic correlation. Furthermore, after substituting DEG-targeting-microRNA for DEG members of each gene set, two gene sets with the microRNA expression profiles were obtained that had prognostic potential.

Conclusions: The malignancy of HCC could be represented by gene sets, and pericancerous liver exhibits important characteristics of liver cancer. The expression level of gene sets not only in HCC but also in the pericancerous liver showed potential for prognosis implying an option for HCC prognosis at an early stage. Additionally, the gene-targeting-microRNA expression profiles also showed prognostic potential, demonstrating that the multi-factorial molecular pathogenesis of HCC is contributed by various genes and microRNAs.

Background

Hepatocellular carcinoma (HCC), is the sixth most prevalent cancer and the third most frequent cause of cancer-related death [1]. More than 50% of the world's HCC cases occur in China (age-standardized incidence rate: men, 35.2/100 000; women, 13.3/100 000) [2]. The pathogenesis of HCC has been reported to be multi-factorial

[3,4]. Liver cirrhosis is the most important risk factor for HCC [1], which occurs in 80%-90% of HCC patients [5]. In China, chronic hepatitis B virus (HBV) infection is another major risk factor [6], which occurs in approximately 85% of HCC patients [7]. Additionally, the great majority of HBV-infected HCC patients (70% and 90%) have coexisting cirrhosis [2].

The complex process of molecular pathogenesis in HCC also indicates that it is caused by multiple types of genes during its development and progression. For years, the combination of microarray and bioinformatics analytical tools have been widely used to find differentially expressed genes in hepatocellular carcinoma and to find differential

* Correspondence: xhhe@shsci.org; xielu@sabit.org

²State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200032, P. R. China

³Shanghai Center for Bioinformation Technology, Shanghai Academy of Science and Technology, Shanghai 201203, P. R. China

Full list of author information is available at the end of the article

diagnostic and prognostic markers [8-15]. Many such studies have used pericancerous liver tissue (assumed to be normal) as the control when selecting differentially changed genes in HCC [8-13]. However, because most pericancerous tissue of HCC is cirrhotic, this assumption could miss important basal molecular changes in the cancer microenvironment. Scientists also attempted to look for differentially expressed genes for prognosis in cirrhosis [15] and non-cancerous liver tissues [14]. As we and other researchers have discovered, dynamic dysregulation exists in the development from cirrhosis to HCC [16], and differentially expressed microRNA in peri-cancer has been used for the prognosis of HCC patients [17].

The low survival rate of HCC patients is largely attributed to the high metastasis rate of HCC. Early studies showed that molecular changes in primary HCC tissue already implied future distant metastasis potential [13]. Additionally, the metastases were reported to be influenced by liver microenvironment that can be represented by inflammation/immune response-related signatures of differentially expressed genes [14]. It would be very interesting to know what kind of molecular changes in the pericancerous tissue of HCC also bear a prediction potential for survival.

In this work, by applying gene expression profiling in hepatocellular carcinoma and pericancerous hepatocellular tissues from HCC patients and in normal liver tissues from healthy individuals, we made an effort to investigate the functional transition in pericancerous liver and cancer liver in HCC patients. We identified expression-changed genes in pericancerous liver and HCC tissue. Then, we conducted functional enrichment analyses to demonstrate the mechanism causing these transitional molecular changes. Additionally, we checked the relationship between the expression level of differentially expressed members of each gene set and the postoperative survival time of HCC patients. We found nine gene sets to be potential prognostic markers. Furthermore, according to the targeting relationships between genes and microRNAs, we also substituted microRNAs for the gene members of each gene set, and we attempted to predict the prognosis with the expression level of the microRNAs that target differentially expressed members of gene sets. Two prognosis-related microRNA sets were identified.

Methods

Ethics statement

All human materials were obtained according to consent regulation and approved by the Ethical Review Committee of the World Health Organization Collaborating Center for Human Products Research (authorized by Shanghai Municipal Government). The individuals in this manuscript have given written informed consent to publish these case details.

Expression profile of mRNA and microRNA

The expression profiling of mRNA and microRNA were performed on three types of liver tissues: HCC, pericancerous liver and normal liver. Forty-five pairs of homogenous human primary hepatocellular carcinoma and adjacent pericancerous liver tissues were collected from the surgical specimen archives of the Department of Pathology, First Affiliated Hospital of Zhejiang University (Hangzhou City, Zhejiang Province, China) and Qidong Liver Cancer Institute (Qidong City, Jiangsu Province, China). The pericancerous liver tissues were collected three centimeters away from any liver tumor. Phenotypic information was collected from patients' records (Additional file 1). And none of the HCC patients had received chemotherapy prior to surgical operation. Ten normal liver tissues were obtained from persons who died in traffic accidents. All of these tissues were freshly frozen at -80°C and confirmed by a pathologist. In each tissue, the total RNA was extracted by TRIzol reagent (Invitrogen, CA, USA); the gene expression was profiled by CapitalBio Human 22k oligonucleotide microarray ([GEO:GPL5918]); and the microRNA expression was profiled by CapitalBio Mammalian miRNA Array Services V1.0 ([GEO:GPL6542]). The expression profiling by array is deposited in Gene Expression Omnibus (GEO) [18] with the accession numbers [GEO:GSE45114] (mRNA) and [GEO:GSE10694] (microRNA) [17].

Differentially expressed genes

Differentially expressed genes (DEG) involved in three comparisons (Cancer/Normal, Cancer/Pericancerous and Pericancerous/Normal) were detected by the limma [19,20] package in Bioconductor [21] with absolute \log_2 -fold-change > 2 and adjusted p-value < 0.001 , which was adjusted by Benjamini and Hochberg's method (BH) [22]. These three groups of DEGs (C/N_all, C/P_all and P/N_all) were further separated into smaller groups, up-regulated DEGs and down-regulated DEGs: C/N_up and C/N_down; C/P_up and C/P_down; and P/N_up and P/N_down.

Gene set enrichment analysis

Gene set enrichment analysis for each group of DEGs was performed by the HTSanalyzer [23] package in Bioconductor with the collection of annotated gene sets provided by the Molecular Signatures Database [24] (MSigDB v4.0, released Jun 7, 2013, including 10295 records). The MSigDB collects various types of gene set, including seven major collections: c1, chromosome and cytogenetic band; c2, online pathway database, publications in PubMed, and knowledge of domain experts, its CP sub-collection collected 1320 Canonical pathways derived from the pathway databases of BioCarta [25], KEGG [26], PID [27], Reactome [28] and four others (SigmaAldrich [29], Signaling Gateway [30], Signal Transduction KE [31], SuperArray [32]); c3, conserved

cis-regulatory motifs, its TFT sub-collection collected 615 gene sets that contain genes sharing a transcription factor binding site defined in the TRANSFAC (version 7.4) database; c4, computational gene sets defined by mining large collections of cancer-oriented microarray data; c5, gene ontology, collected 1454 gene sets derived from the controlled vocabulary of the Gene Ontology (GO) project [33]; c6, oncogenic signatures; and c7, immunologic signatures. Only when the BH-adjusted p-values from a hypergeometric test and Gene Set Enrichment Analysis (GSEA) [24] were both lower than 0.05 was the gene set thought to be significantly enriched with this group of DEGs.

MicroRNAs that target differentially expressed genes

By the RmiR [34] package in Bioconductor, we obtained the targeting relationships between microRNAs and genes that appear in at least three microRNA target databases from six: miRBase [35], TargetScan [36], miRanda [37], tarBase [38], mirTarget2 [39] and PicTar [40]. Then, we obtained the set of microRNAs that target differentially expressed genes in each gene set.

Association between gene (or microRNA) expression profile and postoperative survival time

We used either the DEGs in each enriched gene set or the microRNAs that target DEGs in each enriched gene set to comprise a candidate classifier for prognosis. The associations between gene (or microRNA) expression and postoperative survival time were tested by the phenoTest [41] package in Bioconductor. The effects of the gene expression (or microRNA expression) on survival were tested via the Cox proportional hazards model [42] and Kaplan-Meier estimator [43]. Additionally, these associations were validated on two independent data sets: [GEO:GSE14520] [44,45] (including gene expression profiles of 227 pairs of cancer and pericancerous liver samples, as well as 2 normal liver samples), and the liver hepatocellular carcinoma tumor type from The Cancer Genome Atlas [46] (TCGA LIHC) (including gene and microRNA expression profiled with RNASeq from 27 pairs of cancer and pericancerous liver tissues). The phenotypic information of 227 patients from [GEO:GSE14520] and 27 patients from TCGA LIHC are provided in Additional file 1.

Results

Differentially expressed genes

With the threshold of absolute log₂-fold-change > 2 and adjusted p-value < 0.001, totally 551 differentially expressed genes (DEG) were identified from three comparisons, Cancer/Normal (C/N, 479 DEGs), Cancer/Pericancerous (C/P, 234 DEGs) and Pericancerous/Normal (P/N, 76 DEGs) (Additional file 2). And subgroups of DEGs from each comparison were selected with up or down regulation of DEGs (Figure 1). In Figure 1, sum of the “up-regulated

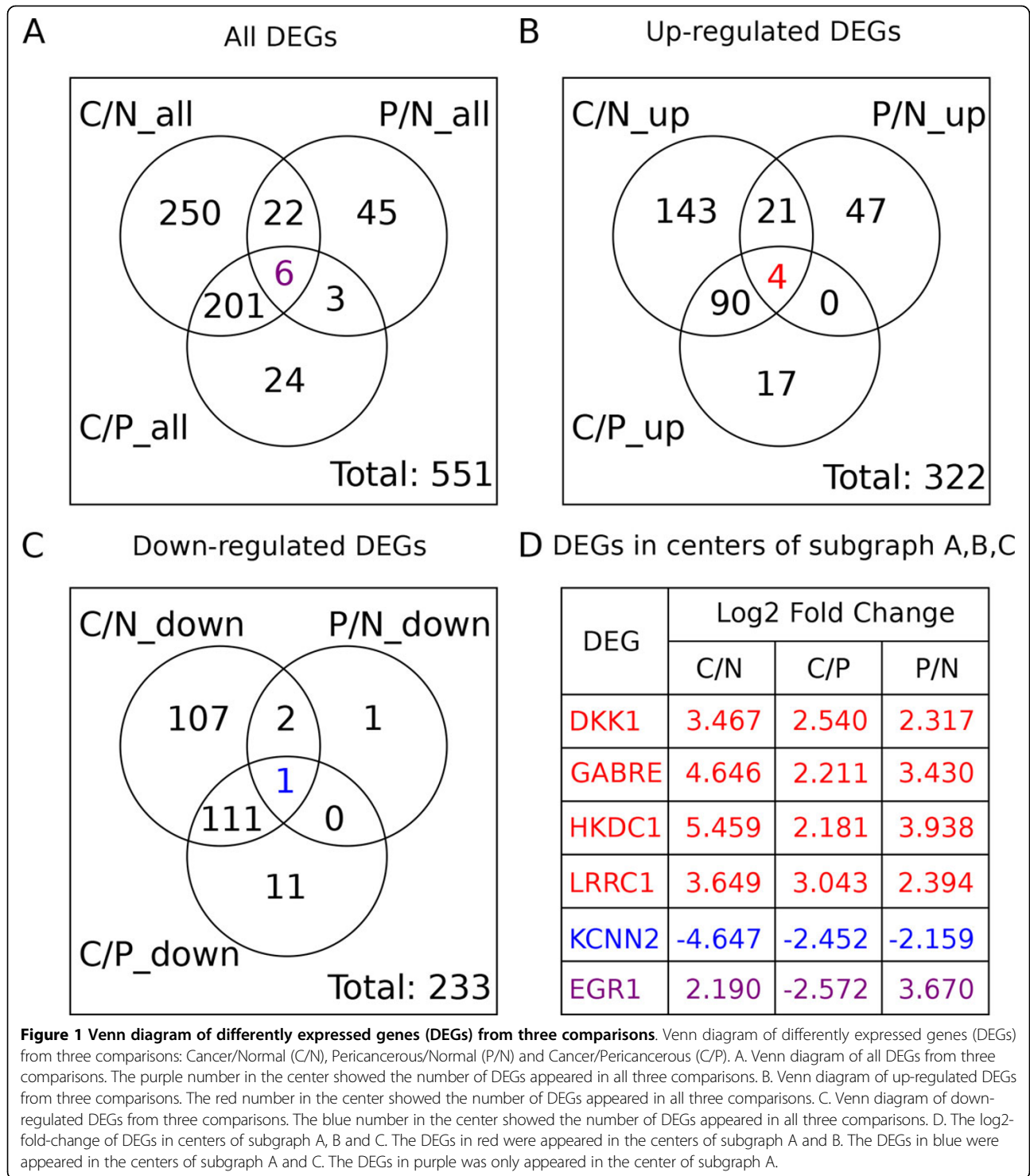
DEGs” (322) and “down-regulated DEGs” (233) are more than “all DEGs” (555 vs. 551), because some genes were up-regulated in one comparison but down-regulated in another, such as *EGR1* listed in Figure 1D, and they appeared in both Figure 1B and 1C.

Among the 551 DEGs, six genes were differentially expressed in all three comparisons (Figure 1). From Figure 1D, *DKK1*, *GABRE*, *HKDC1* and *LRRC1* were up-regulated in pericancerous liver and more up-regulated in cancer liver. The *DKK1* is a Wnt pathway inhibitor, promoting invasion and metastasis of HCC [47], and a serum biomarker for HCC diagnosis [48]. Although the other three DEGs have not been reported in HCC, they are disease related. *GABRE* is related to migraine susceptibility [49]. *HKDC1* is related to Alzheimer disease [50]. And *LRRC1* is DNA repair related [51]. We think they may be important in HCC carcinogenesis. On the contrary, *KCNN2* was down-regulated in pericancerous and more down-regulated in cancer liver. Since *KCNN2* is important for mediating the increase of transepithelial secretion in biliary epithelial cells and prominently expressed in intact liver [52], it seems some function of normal liver was gradually suppressed in pericancerous and cancer liver. And *EGR1* was more up-regulated in pericancerous liver but less up-regulated in cancer liver. Considering that *EGR1* is required for differentiation and mitogenesis [53], the cell proliferation might be up-regulated in both HCC and pericancerous liver, while differentiation might be kept in pericancerous liver but suppressed in HCC.

Gene sets enriched with differentially expressed genes

Gene set enrichment analysis was performed to identify DEG-related functional gene sets. For each subgroup of DEGs in Figure 1, the gene set enrichment analysis (by hypergeometric test and GSEA) was run on 10295 annotated gene sets in MSigDB v4.0, and a small part of them were enriched with the nine subgroups of DEGs (see the nine circles in Figure 1 A, B and C). The intersections of gene sets enriched with different groups of DEGs were counted in a Venn diagram (Additional file 3). Most gene sets were enriched with both C/N DEGs and C/P DEGs. Especially, the gene sets enriched with both C/N_{up} DEGs and C/P_{up} DEGs (or both C/N_{down} DEGs and C/P_{down} DEGs) showed the characters present in pericancerous liver but more dys-regulated in HCC. Thus they would provide us some clues about the gradual carcinogenesis of liver tissue.

We further focused on detailed functional analyses of gene sets enriched in three categories of MSigDB v4.0 collection: c5, Gene Ontology (GO) sets; c2, Canonical pathway sets; and c3, transcription factor targets gene sets (TFT). There are 19 GO terms enriched with both C/N_{up} and C/P_{up} DEGs (Additional file 4, 5), including



biological process (BP) related to “cell cycle” and “mitosis”, as well as cellular component (CC) related to “chromosome” and “spindle”, showing us the character of cell proliferation that is closely related to carcinogenesis. Meanwhile, 21 GO terms were enriched with both C/N_down and C/P_down DEGs (Additional file 4, 5),

including various “metabolism” related BP, CC and MF (molecular function), indicating that metabolisms were disturbed in pericancerous liver and more so in HCC.

Similarly, there are 24 pathways that were enriched with both C/N_up and C/P_up DEGs (Additional file 4, 5). Keywords such as “Cell Cycle”, “G1”, “S”, “G2”, “M” and

“Replication” indicate the genome instability and cell proliferation hallmark of cancer cells [54] being activated. The “p53” and “p73” related pathways indicate DNA damage and apoptosis found in tumorigenesis. At the same time, the *ATR* (ataxia telangiectasia and Rad3-related [55]) pathway, *PLK1* (polo-like kinase 1 [56]) pathway and the Fanconi anemia pathway showed the ability to repair DNA damage in cancer cells. Thus, as a hallmark of HCC, cell proliferation is the result of rebalancing between active apoptosis by DNA damage and active survival by DNA damage repair. Twenty-one pathways were enriched with both C/N_down and C/P_down DEGs (Additional file 4, 5). The most repetitive keywords are “Metabolism” and “PID_HNF3BPATHWAY” (transcription factor network of *FOXA2* and *FOXA3*), hinting that the function of metabolism regulation and the potential for differentiation were abnormal in HCC, because *FOXA2* (forkhead box A2 [57]) and *FOXA3* (forkhead box A3 [58]) are hepatocyte nuclear factors that act as transcriptional activators for liver-specific genes such as albumin and transthyretin. Similar results have been found in mice [59].

Not only GO and pathway gene sets, but the transcription factor targets gene sets (TFTs) also provided functional annotations for DEGs. We found 19 TFTs were enriched with both C/N_up and C/P_up DEGs (Additional file 4, 5), with the cell cycle controlling transcription factor E2F family being the most conspicuous factor. And *E2F3* and *E2F8* were over-expressed in HCC indeed (Additional file 2). At the same time, only one TFT “RGTTAMW-NATT_V\$HNF1_01” was enriched with both C/N_down and C/P_down DEGs (Additional file 4, 5).

From the gene ontology, pathway and transcription factor targets related gene sets enriched with both C/N DEGs and C/P DEGs, we found that during cancer progress of HCC, cell proliferation was gradually up-regulated while metabolism was progressively down-regulated. It is rare to observe such phenomena with direct proofs, the advantage stem from our gene expression profiling of gradually changing samples: from normal, to pericancerous, to cancerous liver tissues.

Association between gene expression profile and postoperative survival time

It is understandable that transitional molecular changes represented by gene sets may demonstrate mechanistic trend of development from normal tissue to cancer tissue, however, whether such changes can be prognostic may be another question.

The DEGs in each enriched gene set might comprise a candidate gene classifier for prognosis. We tested the association between the expression of these candidate gene classifiers and postoperative survival time in our data set, which was 45 HCC patients from [GEO:GSE45114]. Nine gene sets with the expression level of DEGs that associated

with the postoperative survival time in our dataset were also validated in [GEO:GSE14520] (227 HCC patients) (Table 1). As shown in Table 1, Figure 2, 3 and Additional file 6, the expression profile of sets of DEGs in HCC, even pericancerous liver could be used for prognosis.

The first three gene sets in Table 1 showed prognosis potential with up-regulated DEGs in cancer liver. Their DEG members expression level in cancer could be used for prognosis in both our 45 HCC patients from [GEO:GSE45114] and the 227 HCC patients from [GEO:GSE14520] ($P < 0.05$ and $HR > 0$). The positive HR (hazard ratio) means the higher DEGs expression the worse the prognosis. In Figure 2, we show the prognosis ability of nine DEGs in gene set “chr1q32” which was reported to be the most recurrently gained genomic region in HCC [60]. Another gene set “KAUFFMANN_MELANOMA_RELAPSE_UP” [61] contains DNA repair and replication related genes (Additional file 6).

The next three gene sets in Table 1 showed prognosis potential of pericancerous liver with up-regulated DEGs. Gene set “BROWNE_HCMV_INFECTION_2HR_UP” contains genes that were related to hepatic inflammation and cirrhosis [62]. Their expression level may represent not only inflammation and cirrhosis but also carcinogenesis of HCC (Figure 3). And the gene set “ENK_UV_RESPONSE_EPIDERMIS_DN” [63] contains genes related to DNA damage repair (Additional file 6).

Besides up-regulated DEGs, the down-regulated DEGs in cancer liver also showed prognosis potential in the last three gene sets (Table 1 and Additional file 6). Here, negative HR (hazard ratio) means the lower DEG expression the worse the prognosis.

In summary, prognosis of HCC patients could be predicted with expression profiles of both up-regulated DEGs and down-regulated DEGs enriched in certain functional gene sets.

Association between microRNA expression profile and postoperative survival time

Gene sets enriched with DEGs either in C/N, C/P or P/N were shown to have prognosis potential, as reported above. MicroRNA profiling data is also available for the 45 HCC patients with paired pericancer/cancer samples. Since microRNA expression signatures in hepatocellular carcinoma have been stated to possess prognostic value before [17,64], we would like to see in our work, whether DEGs related microRNA sets could be prognostic. We identified the targeting relationships between microRNAs and genes that appear in at least three microRNA target databases from six: miRBase [35], TargetScan [36], miRanda [37], tarBase [38], mirTarget2 [39] and PicTar [40]. The microRNAs that target DEGs in each enriched gene set comprise a candidate microRNA set for prognosis prediction. Then we tested the association between the

Table 1 Gene sets associated to postoperative survival time with DEG expression profile (validated in [GEO:GSE14520]).

Gene set	Enriched with DEG group	DEG Count	DEG	Profile*
chr1q32	C/N_up	9	BATF3, C1orf106, CENPF, MDM4, NEK2, RAB1F, TLR5, TRAF5, UBE2T	C
KAUFFMANN_MELANOMA_RELAPSE_UP	C/N_up	12	CENPF, CHEK1, CHEK2, FANCD2, GINS2, MCM6, MSH2, RAD54L, RFC4, RRM2, SMC2, TOP2A	C
PETROVA_PROX1_TARGETS_UP	C/N_up & C/P_up	5	BUB1B, CCNE2, CDK1, MCM6, TOP2A	C & C/P
BROWNE_HCMV_INFECTION_2HR_UP	P/N_up	4	EGR1, FOS, NR4A2, NR4A3	P/N
ENK_UV_RESPONSE_EPIDERMIS_DN	P/N_up	10	AREG, CH25H, EGR1, EGR2, FJX1, GSN, LXN, NR4A2, PTGS2, STK17B	P/N
GSE9988_LOW_LPS_VS_CTRL_TREATED_MONOCYTE_UP	P/N_up	8	AREG, EGR1, IL6, NR4A2, NR4A3, PTGS2, RGS1, SERPINB8	P/N
MODULE_43	C/N_down	17	ACADL, ACADSB, CAT, CYP1A1, CYP26A1, CYP2C8, CYP3A4, CYP3A7, CYP4A11, DAO, ETFDH, FMO3, GYS2, LEPR, PPP1R1A, QDPR, TREH	C
MODULE_99	C/N_down	17	ABAT, AFM, BBOX1, BCHE, CRHBP, F11, FETUB, IGF1, LEPR, LIFR, MARCO, NAT2, OTC, PCK1, SLC10A1, SLC22A1, UGT8	C/N
PKCA_DN.V1_UP	C/N_down	17	ADRA1A, BBOX1, CYP2C18, CYP2C8, DIO1, FBP1, LECT2, NR1B3, PCK1, PIPOX, PRODH2, RDH16, RDH5, SEC14L2, SLC2A2, TREH, UPB1	C

*"Profile": the expression profile used for prognosis; "C" means the absolute expression value of genes in cancer; while "C/P" and "P/N" mean the relative gene expression level of cancer/pericancerous liver and pericancerous/normal liver respectively.

expression of these microRNAs and postoperative survival times in our 45 patients from [GEO:GSE10694]. Two prognostic microRNA gene sets were validated in an independent test dataset TCGA LIHC (27 HCC patients with RNASeq data) (Table 2).

Gene set "SMID_BREAST_CANCER_BASAL_DN" contains genes that are down-regulated in basal subtype of breast cancer samples [65]. We found that 32 member genes were down-regulated in HCC relative to normal liver and nine of them were targeted by 37 microRNAs. The 37 microRNAs expression profile in cancer liver could be used for prognosis (Figure 4). The positive HR (hazard ratio) means the higher expression the worse prognosis.

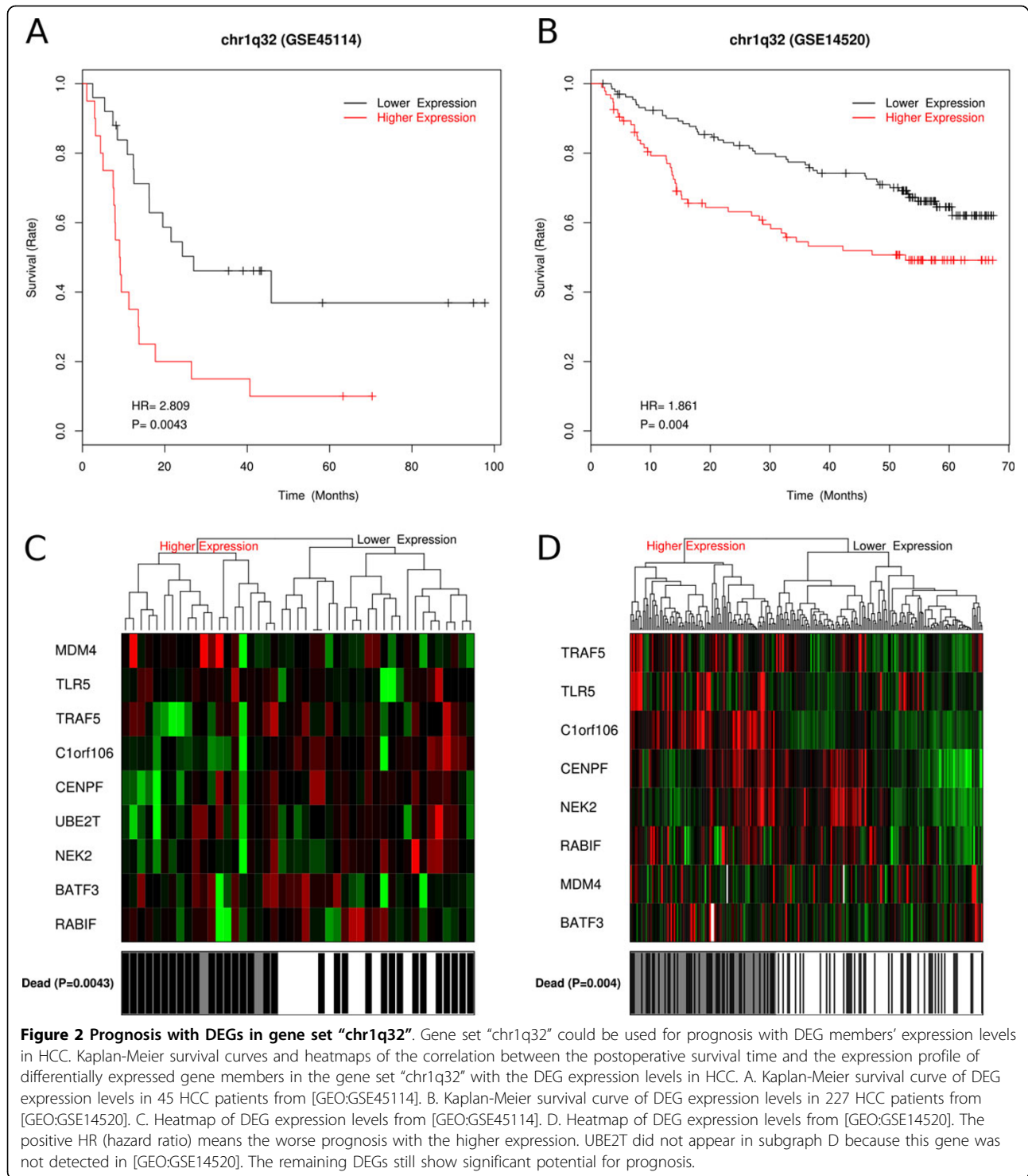
The other gene set "SMID_BREAST_CANCER_LUMINAL_B_UP" contains genes that are up-regulated in the luminal B subtype of breast cancer [65]. Its 12 member genes were down-regulated in HCC relative to normal liver and four of them were targeted by 22 microRNAs. The 22 microRNAs expression profile in the cancer liver could be used for prognosis (Additional file 6). Interestingly, the four DEGs are subset of the nine DEGs mentioned in above gene set (Table 2), which shows the similarity and difference between basal subtype and luminal subtype of breast cancers.

Most of the microRNAs listed in Table 2 have been annotated to be related to HCC in The human microRNA disease database (HMDD) [66] (Additional file 7). Such as the cell proliferation related microRNAs: hsa-mir-18a, hsa-mir-93, hsa-mir-96; and cancer recurrence

related microRNAs: hsa-mir-148a, hsa-mir-18a, hsa-mir-18b, hsa-mir-19a, hsa-mir-22, hsa-mir-221, hsa-mir-222, hsa-mir-96. In Table 2 there are some microRNAs have not been recorded to be HCC related by HMDD, including: hsa-miR-136, hsa-miR-206, hsa-miR-26b, hsa-miR-302a, hsa-miR-302d, hsa-miR-340, hsa-miR-410, hsa-miR-488, hsa-miR-495, hsa-miR-506. They may be potentially HCC related.

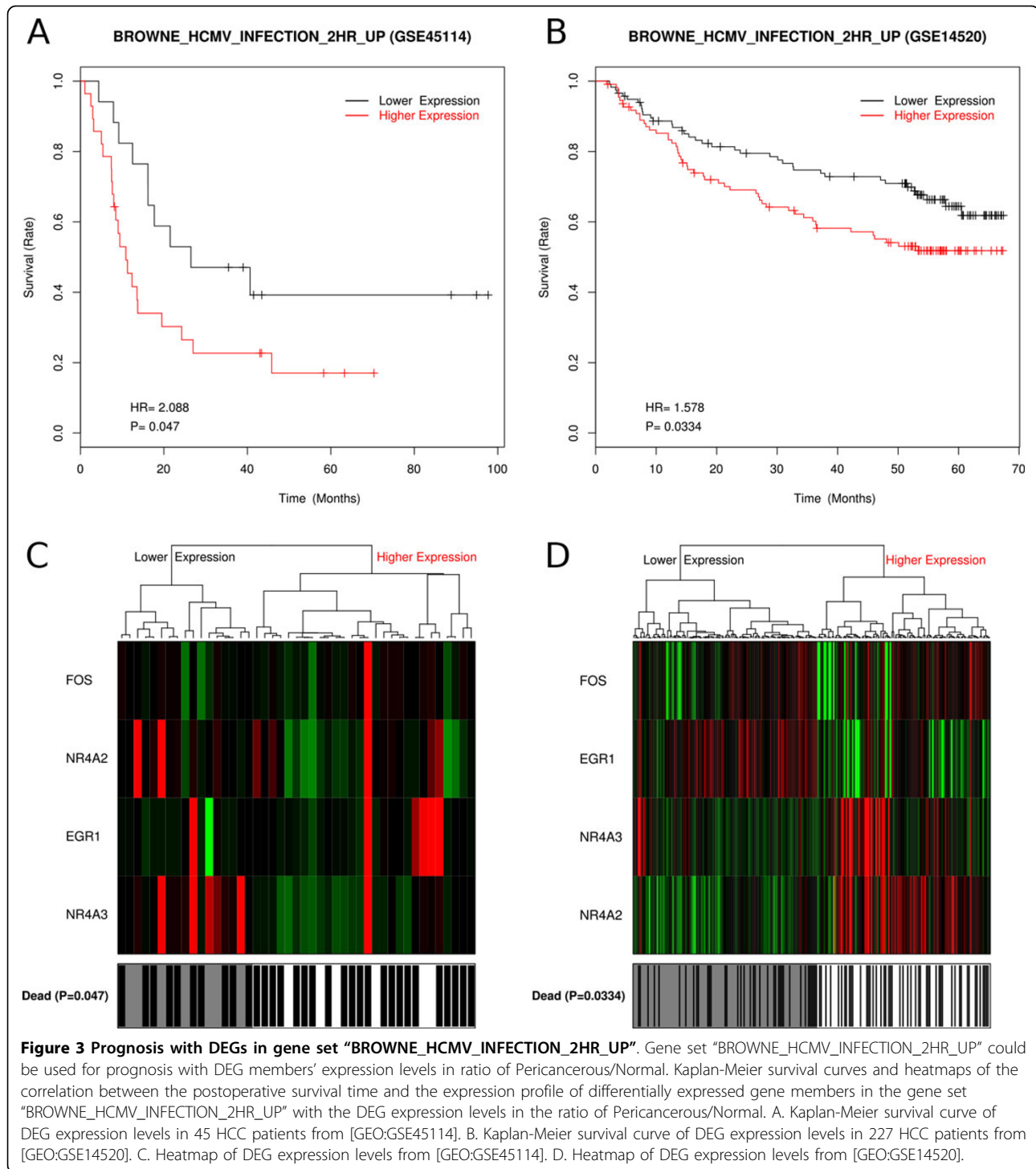
Discussion

There have been numerous studies of hepatocellular carcinoma(HCC) in comparison with pericancerous tissue as normal control, in the purpose of identifying differentially expressed genes, modules, networks etc., in order to find cancer biomarkers, cluster samples, or to predict prognosis. Such studies especially on Chinese HCC patient samples take on a strong assumption that pericancerous liver tissue of HCC is normal, while this in a large percentage is wrong. Most patients diagnosed with HCC in China already went through years of liver cirrhotic change because of chronic HBV infection, alcoholism, or fatty liver etc. Therefore, in this work of ours, we designed a set of normal liver tissues as control. With such a design, we were able to identify differentially expressed genes (DEGs) with a gradual up-regulation from normal to pericancerous to cancerous liver, or DEGs with a gradual down-regulation. Further gene set enrichment analysis (GSEA) on GO terms, pathway, and transcription factor targets suggested the main up-regulated trend to be in cell cycle and proliferation, and the main down-regulated trend to



be metabolism. Although such conclusions may not be totally novel, it is nice to see such direct proof of gradual molecular transitions in liver carcinogenesis. More in-depth detailed analyses of the gradually changed gene sets may even lead to clues for early diagnosis, however it is beyond the scope of this paper's discussion.

Instead, we made efforts to testify whether gene sets enriched with gradually changing DEGs have prognostic value. Many previous researches proposed lists of DEGs, pathways, network modules (the latter two can be considered as gene sets) to predict prognosis for HCC patients. We used somewhat a combined approach. Instead of



using groups of single DEGs that would lack functional binding, or full gene sets that would contain too many genes, we used DEGs grouped in preselected enriched gene sets as classifiers. The advantage is that the classifier is relatively small, and the DEGs share a common gene function family. Indeed we were able to identify nine such

gene set DEGs classifiers possessing prognostic prediction power, and could even be validated in an independent dataset with larger patient number. Quite a few such gene sets behold cell proliferation or DNA repair functions in liver cancer tissues, or inflammation function in pericancerous liver tissues.

Table 2 Gene sets associated to postoperative survival time with microRNA expression profile (validated in TCGA LIHC).

Gene set *	DEG	microRNAs targeting to DEG	Count
SMID_BREAST_CANCER_BASAL_DN	ABAT ASPA CXCL14 ESR1 FAM134B IGF1 PBLD PCK1 SORD	hsa-let-7b, hsa-miR-130a, hsa-miR-130b, hsa-miR-136, hsa-miR-148a, hsa-miR-148b, hsa-miR-152, hsa-miR-181c, hsa-miR-181d, hsa-miR-182, hsa-miR-18a, hsa-miR-18b, hsa-miR-19a, hsa-miR-206, hsa-miR-22, hsa-miR-221, hsa-miR-222, hsa-miR-26b, hsa-miR-29a, hsa-miR-29c, hsa-miR-302a, hsa-miR-302b, hsa-miR-302c, hsa-miR-302d, hsa-miR-31, hsa-miR-340, hsa-miR-372, hsa-miR-373, hsa-miR-410, hsa-miR-425, hsa-miR-488, hsa-miR-495, hsa-miR-506, hsa-miR-520b, hsa-miR-520e, hsa-miR-93, hsa-miR-96	9 Genes 37 miR
SMID_BREAST_CANCER_LUMINAL_B_UP	ABAT ESR1 FAM134B SORD	hsa-miR-130a, hsa-miR-130b, hsa-miR-148a, hsa-miR-148b, hsa-miR-152, hsa-miR-181c, hsa-miR-181d, hsa-miR-182, hsa-miR-18a, hsa-miR-18b, hsa-miR-19a, hsa-miR-22, hsa-miR-221, hsa-miR-222, hsa-miR-302a, hsa-miR-302c, hsa-miR-302d, hsa-miR-31, hsa-miR-372, hsa-miR-506, hsa-miR-93, hsa-miR-96	4 Genes 22 miRs

*These two gene sets were both enriched with C/N_down group DEGs. And the absolute expression value of microRNA in cancer, which detected by array in [GEO:GSE10694] or by RNASeq in TCGA LIHC were used for prognosis.

MicroRNA (miRNA) as a new kind of regulatory biomarker, has been investigated in many cancers in recent years. In our previous works, individual miRNAs and miRNA regulatory network modules have been successfully applied in HCC prognosis prediction [17,67,68]. In this work, we took a simple approach. Since some of the gene sets enriched with gradually changing DEGs in liver carcinogenesis have been proved to possess prognosis potential, we substituted such gene sets with miRNAs targeting the DEGs they contained. To ensure the substitutions are relevant all miRNA-DEG target relationships must be carefully curated from multiple databases and prediction algorithms. Two gene sets substituted with miRNAs acquired prognostic power, and could be validated in a TCGA RNASeq dataset which has miRNA expression data of paired HCC samples available. This may actually represent a simple approach to quickly discover relevant miRNAs which might have caused the dysregulation of the DEGs that are associated with prognosis. Traditionally differentially expressed miRNAs should first be detected and secondly correlated to their downstream targets and further to functional applications.

Figure 1 and Additional file 3 indicated the similarities between pericancerous and normal liver, when comparing to HCC. This proves the rationality for many researchers who take pericancerous tissues as control. Similarly, researchers found that gene expression pattern is more significantly related to physiological condition rather than tissue spatial distance [69]. They reported that different cancer tissues may show common gene expression patterns. Our results might provide an evidence for that: some prognosis biomarkers we found in HCC also play important roles in other cancers, such as melanoma and breast cancer (Additional file 6 and Figure 4). At the same time, we found that pericancerous liver shared some characters of HCC, which provided the probability for

prognosis prediction with gene expression profiles of pericancerous liver (Figure 3 and Additional file 6).

There are of course limitations to our work. The patient sample size is not big, and the normal samples are from healthy individuals who died accidentally, rather than real normal liver sample of the same HCC patient, which is hardly possible to get. Therefore the gradual changes from normal to pericancerous to cancerous liver tissues observed in this dataset may not be very steady accessible features that can be easily applied clinically. However our strategy does put an emphasis on the importance to study the cirrhotic and inflammatory nature of pericancerous tissue in HCC patients, which show both carcinogenesis trend and prognostic potential. In the future, integrating sequence information from DNaseq and RNASeq as well as clinical information in bigger sample size data sets may benefit such purpose.

Conclusions

In this work, Based on differentially expressed genes (DEGs) detected from normal, pericancerous, cancerous liver samples by array technology, and the annotated gene sets from GSEA MSigDB, we managed to show some molecular transitional changes represented by different GO, pathway, regulatory gene sets. DEGs profile of nine of such gene sets could be applied to predict hepatocellular carcinoma (HCC) patient survival. Two gene sets acquired prognostic capacity after being substituted with microRNAs targeting the DEGs contained in the original gene set. Both gene set prognosis and miRNA set prognosis were validated with independent HCC patients gene expression or RNASeq dataset. Our work represents an effort to study pericancerous nature of HCC, and a simple way to identify relevant regulatory miRNAs to DEGs.

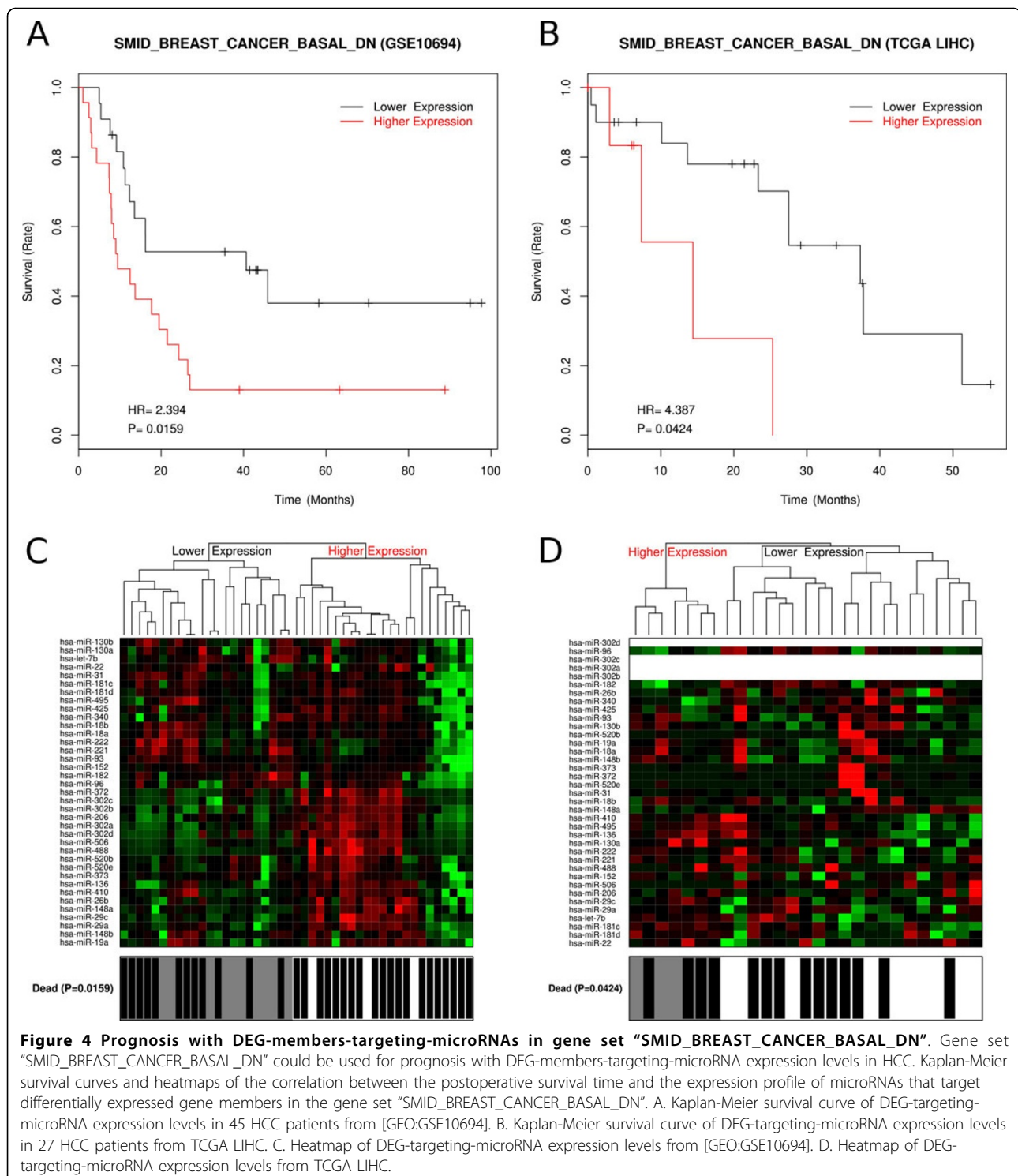


Figure 4 Prognosis with DEG-members-targeting-microRNAs in gene set "SMID_BREAST_CANCER_BASAL_DN". Gene set "SMID_BREAST_CANCER_BASAL_DN" could be used for prognosis with DEG-members-targeting-microRNA expression levels in HCC. Kaplan-Meier survival curves and heatmaps of the correlation between the postoperative survival time and the expression profile of microRNAs that target differentially expressed gene members in the gene set "SMID_BREAST_CANCER_BASAL_DN". A. Kaplan-Meier survival curve of DEG-targeting-microRNA expression levels in 45 HCC patients from [GEO:GSE10694]. B. Kaplan-Meier survival curve of DEG-targeting-microRNA expression levels in 27 HCC patients from TCGA LIHC. C. Heatmap of DEG-targeting-microRNA expression levels from [GEO:GSE10694]. D. Heatmap of DEG-targeting-microRNA expression levels from TCGA LIHC.

Additional material

Additional file 1: Tables, phenotypic data of HCC patients.

Phenotypic data of: 45 patients from [GEO:GSE45114] and [GEO:GSE10694]; 227 patients from [GEO:GSE14520]; 27 patients from TCGA LIHC.

Additional file 2: Tables, DEGs and gene sets enriched with DEGs.

The 551 identified differentially expressed genes (DEGs) of three comparisons: Cancer/Normal (C/N), Cancer/Pericancerous (C/P) and Pericancerous/Normal (P/N). The value "NA" means that this gene (row head) is not a DEG in this comparison (column head). And the 868 non repetitive gene sets that enriched with the nine groups of DEGs (nine circles in Figure 1) by both two enrich method (hypergeometric test and GSEA). The value "NA" means that this gene set (row head) is not enriched with this group of DEGs (column head).

Additional file 3: Figure, venn diagram of gene sets enriched with DEGs from three comparisons.

Venn diagram of gene sets enriched with DEGs from three comparisons: Cancer/Normal (C/N), Pericancerous/Normal (P/N) and Cancer/Pericancerous (C/P). A. Venn diagram of gene sets enriched with the all DEGs from three comparisons. B. Venn diagram of gene sets enriched with the up-regulated DEGs from three comparisons. The red number showed the number of gene sets enriched with both C/N_{up} DEGs and C/P_{up} DEGs. C. Venn diagram of gene sets enriched with the down-regulated DEGs from three comparisons. The blue number showed the number of gene sets enriched with both C/N_{down} DEGs and C/P_{down} DEGs. D. Counts of gene ontology, pathway and transcription factor targets gene sets enriched with both C/N DEGs and C/P DEGs. The numbers in red were covered by red number in subgraph B. The numbers in blue were covered by blue number in subgraph C.

Additional file 4: Figures, venn diagram of GOs, Pathways and TFTs enriched with DEGs from three comparisons.

Venn diagram of gene sets about gene ontology terms (GO), Pathways and Transcription factor targets (TFT) enriched with DEGs from three comparisons: Cancer/Normal (C/N), Pericancerous/Normal (P/N) and Cancer/Pericancerous (C/P). A. Venn diagram of gene sets enriched with the all DEGs from three comparisons. B. Venn diagram of gene sets enriched with the up-regulated DEGs from three comparisons. The red number showed the number of gene sets enriched with both C/N_{up} DEGs and C/P_{up} DEGs. C. Venn diagram of gene sets enriched with the down-regulated DEGs from three comparisons. The blue number showed the number of gene sets enriched with both C/N_{down} DEGs and C/P_{down} DEGs.

Additional file 5: Tables, gene sets enriched with both C/N DEGs and C/P DEGs.

Gene sets about Gene ontology terms (GOs), Pathways and Transcription factor targets gene sets (TFTs) enriched with both C/N DEGs and C/P DEGs.

Additional file 6: Figures, gene sets used for prognosis with expression profile of DEG members.

Kaplan-Meier survival curves and heatmaps of the correlation between the postoperative survival time and the expression profile of DEG members in the gene set "KAUFFMANN_MELANOMA_RELAPSE_UP", "PETROVA_PROX1_TARGETS_UP", "ENK_UV_RESPONSE_EPIDERMIS_DN", "GSE9988_LOW_LPS_VS_CTRL_TREATED_MONOCYTE_UP", "MODULE_43", "MODULE_99" and "PKCA_DN.V1_UP". Each figure includes four subgraphs: A. Kaplan-Meier survival curve of DEG expression levels in 45 HCC patients from [GEO:GSE45114]. B. Kaplan-Meier survival curve of DEG expression levels in 227 HCC patients from [GEO:GSE14520]. C. Heatmap of DEG expression levels in 45 HCC patients from [GEO:GSE45114]. D. Heatmap of DEG expression levels in 227 HCC patients from [GEO:GSE14520]. And the Kaplan-Meier survival curves and heatmaps of the correlation between the postoperative survival time and expression profile of DEG-targeting microRNAs in gene set "SMID_BREAST_CANCER_LUMINAL_B_UP" which validated with 27 HCC patients from TCGA LIHC. Including four subgraphs: A. Kaplan-Meier survival curve of DEG expression levels in 45 HCC patients from [GEO:GSE10694]. B. Kaplan-Meier survival curve of DEG expression levels in 27 HCC patients from TCGA LIHC. C. Heatmap of DEG expression levels in 45 HCC patients from [GEO:GSE10694]. D. Heatmap of DEG expression levels in 27 HCC patients from TCGA LIHC. (Note: The positive HR (hazard ratio) means the higher expression the worse prognosis. While

the negative HR (hazard ratio) means the lower expression the worse prognosis. Some genes may not appear in subgraph D, because those genes (or microRNAs) were not detected in [GEO:GSE14520] (or TCGA LIHC). The remaining DEGs (or microRNAs) still show significant potential for prognosis.)

Additional file 7: Table, microRNAs annotation recorded by the human microRNA disease database (HMDD). Most microRNAs listed in Table 2 have been annotated by the human microRNA disease database (HMDD). The table list the 27 HCC related microRNAs with their references and descriptions annotated by HMDD.

List of abbreviations used

HCC: Hepatocellular carcinoma; DEG: Differentially expressed genes; GSEA: Gene set enrichment analysis; GEO: Gene expression omnibus; MSigDB: Molecular signatures database; GO: Gene ontology; TCGA: The cancer genome atlas; LIHC: Liver hepatocellular carcinoma; C/N: Cancer/Normal; C/P: Cancer/Pericancerous; P/N: Pericancerous/Normal; TFT: Transcription factor targets; HR: Hazard ratio; HMDD: Human microRNA disease database.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LW performed the analyses of this research and drafted the manuscript. BL participated in microRNA targeting, and revision of manuscript. YZ participated in gene set enrichment analysis. WL participated in survival analysis. JG designed this work. XH participated in design and revision. LX participated in design and wrote part of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Written consent was obtained from the patient or their relative for publication for this study. This work was supported by grants from the National 973 Key Basic Research Program (2013CB910500), the National Natural Science Foundation of China (81125016, 81071637, 91029728) and the Key Infectious Disease Project [2012ZX10002012-014].

Declarations

Publication of this article was funded by State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiao Tong University School of Medicine.

This article has been published as part of *BMC Genomics* Volume 15 Supplement 1, 2014: Selected articles from the Twelfth Asia Pacific Bioinformatics Conference (APBC 2014): Genomics. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcgenomics/supplements/15/S1>.

Authors' details

¹Shanghai Medical College, Fudan University, Shanghai 200032, P. R. China. ²State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200032, P. R. China. ³Shanghai Center for Bioinformatics Technology, Shanghai Academy of Science and Technology, Shanghai 201203, P. R. China. ⁴School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, P. R. China.

Published: 24 January 2014

References

1. Forner A, Llovet JM, Bruix J: **Hepatocellular carcinoma.** *Lancet* 2012, **379**:1245-1255.
2. El-Serag HB, Rudolph KL: **Hepatocellular carcinoma: epidemiology and molecular carcinogenesis.** *Gastroenterology* 2007, **132**:2557-2576.
3. Teufel A, Staib F, Kanzler S, Weinmann A, Schulze-Bergkamen H, Galle PR: **Genetics of hepatocellular carcinoma.** *World J Gastroenterol WJG* 2007, **13**:2271-2282.
4. Thorgeirsson SS, Grisham JW: **Molecular pathogenesis of human hepatocellular carcinoma.** *Nat Genet* 2002, **31**:339-346.

5. Fattovich G, Stroffolini T, Zagni I, Donato F: **Hepatocellular carcinoma in cirrhosis: incidence and risk factors.** *Gastroenterology* 2004, **127**(5 Suppl 1): S35-50.
6. Kumagi T, Hiasa Y, Hirschfield GM: **Hepatocellular carcinoma for the non-specialist.** *BMJ* 2009, **339**:b5039.
7. Tanaka M, Katayama F, Kato H, Tanaka H, Wang J, Qiao YL, Inoue M: **Hepatitis B and C virus infection and hepatocellular carcinoma in China: a review of epidemiology and control measures.** *J Epidemiol Jpn Epidemiol Assoc* 2011, **21**:401-416.
8. Choi JK, Choi JY, Kim DG, Choi DW, Kim BY, Lee KH, Yeom YI, Yoo HS, Yoo OJ, Kim S: **Integrative analysis of multiple gene expression profiles applied to liver cancer study.** *FEBS Lett* 2004, **565**:93-100.
9. Neo SY, Leow CK, Vega VB, Long PM, Islam AFM, Lai PBS, Liu ET, Ren EC: **Identification of discriminators of hepatoma by gene expression profiling using a minimal dataset approach.** *Hepatol Baltim Md* 2004, **39**:944-953.
10. Tanaka S, Arai S, Yasen M, Mogushi K, Su NT, Zhao C, Imoto I, Eishi Y, Inazawa J, Miki Y, Tanaka H: **Aurora kinase B is a predictive factor for the aggressive recurrence of hepatocellular carcinoma after curative hepatectomy.** *Br J Surg* 2008, **95**:611-619.
11. Wang K, Liu J, Yan ZL, Li J, Shi LH, Cong WM, Xia Y, Zou QF, Xi T, Shen F, Wang HY, Wu MC: **Overexpression of aspartyl-(asparaginyl)-beta-hydroxylase in hepatocellular carcinoma is associated with worse surgical outcome.** *Hepatol Baltim Md* 2010, **52**:164-173.
12. Yang JD, Sun Z, Hu C, Lai J, Dove R, Nakamura I, Lee JS, Thorgeirsson SS, Kang KJ, Chu IS, Roberts LR: **Sulfatase 1 and sulfatase 2 in hepatocellular carcinoma: associated signaling pathways, tumor phenotypes, and survival.** *Genes Chromosomes Cancer* 2011, **50**:122-135.
13. Ye QH, Qin LX, Forgues M, He P, Kim JW, Peng AC, Simon R, Li Y, Robles AI, Chen Y, Ma ZC, Wu ZQ, Ye SL, Liu YK, Tang ZY, Wang XW: **Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning.** *Nat Med* 2003, **9**:416-423.
14. Budhu A, Forgues M, Ye QH, Jia HL, He P, Zanetti KA, Kammula US, Chen Y, Qin LX, Tang ZY, Wang XW: **Prediction of venous metastases, recurrence, and prognosis in hepatocellular carcinoma based on a unique immune response signature of the liver microenvironment.** *Cancer Cell* 2006, **10**:99-111.
15. Paradis V, Bièche I, Dargère D, Laurendeau I, Laurent C, Bioulac Sage P, Degott C, Belghiti J, Vidaud M, Bedossa P: **Molecular profiling of hepatocellular carcinomas (HCC) using a large-scale real-time RT-PCR approach: determination of a molecular diagnostic index.** *Am J Pathol* 2003, **163**:733-741.
16. Huang T, Liu L, Liu Q, Ding G, Tan Y, Tu Z, Li Y, Dai H, Xie L: **The role of Hepatitis C Virus in the dynamic protein interaction networks of hepatocellular cirrhosis and carcinoma.** *Int J Comput Biol Drug Des* 2011, **4**:5-18.
17. Li W, Xie L, He X, Li J, Tu K, Wei L, Wu J, Guo Y, Ma X, Zhang P, Pan Z, Hu X, Zhao Y, Xie H, Jiang G, Chen T, Wang J, Zheng S, Cheng J, Wan D, Yang S, Li Y, Gu J: **Diagnostic and prognostic implications of microRNAs in human hepatocellular carcinoma.** *Int J Cancer* 2008, **123**:1616-1622.
18. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** *Nucleic Acids Res* 2002, **30**:207-210.
19. Smyth GK: **Limma: linear models for microarray data.** *Bioinforma Comput Biol Solut Using R Bioconductor* New York: Springer; 2005, 397-420.
20. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3**:Article3.
21. **Bioconductor.** [http://www.bioconductor.org].
22. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.** *J R Stat Soc Ser B Methodol* 1995, **57**:289-300.
23. Wang X, Terfve C, Rose JC, Markowitz F: **HTSanalyzer: an R/Bioconductor package for integrated network analysis of high-throughput screens.** *Bioinforma Oxf Engl* 2011, **27**:879-880.
24. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci USA* 2005, **102**:15545-15550.
25. **BioCarta.** [http://www.biocarta.com].
26. Kanehisa M, Goto S: **KEGG: Kyoto Encyclopedia of Genes and Genomes.** *Nucleic Acids Res* 2000, **28**:27-30.
27. **PID.** [http://pid.nci.nih.gov].
28. **Reactome.** [http://www.reactome.org].
29. **SigmaAldrich.** [http://www.sigmaaldrich.com/life-science.html].
30. **Signaling Gateway.** [http://www.signaling-gateway.org].
31. **Signal Transduction KE.** [http://stke.sciencemag.org].
32. **SuperArray.** [http://www.superarray.com].
33. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25-29.
34. Francesco F: **RmiR: Package to work with miRNAs and miRNA targets with R. R package version 1.16.0.**
35. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ: **miRBase: tools for microRNA genomics.** *Nucleic Acids Res* 2008, **36**(Database):D154-158.
36. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP: **Weak seed-pairing stability and high target-site abundance decrease the proficiency of Isy-6 and other microRNAs.** *Nat Struct Mol Biol* 2011, **18**:1139-1146.
37. Betel D, Wilson M, Gabow A, Marks DS, Sander C: **The microRNA.org resource: targets and expression.** *Nucleic Acids Res* 2008, **36**(suppl 1): D149-D153.
38. Papadopoulos GL, Reczko M, Simossis VA, Sethupathy P, Hatzigeorgiou AG: **The database of experimentally supported targets: a functional update of TarBase.** *Nucleic Acids Res* 2009, **37**(Database):D155-158.
39. Wang X: **miRDB: a microRNA target prediction and functional annotation database with a wiki interface.** *RNA N Y N* 2008, **14**:1012-1017.
40. **PicTar.** [http://pictar.mdc-berlin.de].
41. Planet E: **phenoTest: Tools to test association between gene expression and phenotype in a way that is efficient, structured, fast and scalable. We also provide tools to do GSEA (Gene set enrichment analysis) and copy number variation.. R package version 1.6.0.** 2010.
42. Cox D: **Regression models and life-tables.** *J R Stat Soc Ser B* 1972, **34**:187-220.
43. Kaplan EL, Meier P: **Nonparametric Estimation from Incomplete Observations.** *J Am Stat Assoc* 1958, **53**:457.
44. 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**:10202-10212.
45. Roessler S, Long EL, Budhu A, Chen Y, Zhao X, Ji J, Walker R, Jia HL, Ye QH, Qin LX, Tang ZY, He P, Hunter KW, Thorgeirsson SS, Meltzer PS, Wang XW: **Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival.** *Gastroenterology* 2012, **142**:957-966.e12.
46. **The Cancer Genome Atlas (TCGA).** [http://cancergenome.nih.gov].
47. Tao YM, Liu Z, Liu HL: **Dickkopf-1 (DKK1) promotes invasion and metastasis of hepatocellular carcinoma.** *Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver* 2013, **45**:251-257.
48. Shen Q, Fan J, Yang XR, Tan Y, Zhao W, Xu Y, Wang N, Niu Y, Wu Z, Zhou J, Qiu SJ, Shi YH, Yu B, Tang N, Chu W, Wang M, Wu J, Zhang Z, Yang S, Gu J, Wang H, Qin W: **Serum DKK1 as a protein biomarker for the diagnosis of hepatocellular carcinoma: a large-scale, multicentre study.** *Lancet Oncol* 2012, **13**:817-826.
49. Fernandez F, Esposito T, Lea RA, Colson NJ, Ciccociola A, Gianfrancesco F, Griffiths LR: **Investigation of gamma-aminobutyric acid (GABA) A receptors genes and migraine susceptibility.** *BMC Med Genet* 2008, **9**:109.
50. Grupe A, Li Y, Rowland C, Nowotny P, Hinrichs AL, Smemo S, Kauwe JSK, Maxwell TJ, Cherny S, Doil L, Tacey K, van Luchene R, Myers A, Wavrant-De Vrièze F, Kaleem M, Hollingworth P, Jehu L, Foy C, Archer N, Hamilton G, Holmans P, Morris CM, Catanese J, Sninsky J, White TJ, Powell J, Hardy J, O'Donovan M, Lovestone S, Jones L, et al: **A scan of chromosome 10 identifies a novel locus showing strong association with late-onset Alzheimer disease.** *Am J Hum Genet* 2006, **78**:78-88.
51. Svendsen JM, Smogorzewska A, Sowa ME, O'Connell BC, Gygi SP, Elledge SJ, Harper JW: **Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair.** *Cell* 2009, **138**:63-77.

52. Feranchak AP, Doctor RB, Troetsch M, Brookman K, Johnson SM, Fitz JG: **Calcium-dependent regulation of secretion in biliary epithelial cells: the role of apamin-sensitive SK channels.** *Gastroenterology* 2004, **127**:903-913.
53. Sukhatme VP, Cao XM, Chang LC, Tsai-Morris CH, Stamenkovich D, Ferreira PC, Cohen DR, Edwards SA, Shows TB, Curran T: **A zinc finger-encoding gene coregulated with c-fos during growth and differentiation, and after cellular depolarization.** *Cell* 1988, **53**:37-43.
54. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation.** *Cell* 2011, **144**:646-674.
55. Cimprich KA, Shin TB, Keith CT, Schreiber SL: **cDNA cloning and gene mapping of a candidate human cell cycle checkpoint protein.** *Proc Natl Acad Sci USA* 1996, **93**:2850-2855.
56. Holtrich U, Wolf G, Bräuninger A, Karn T, Böhme B, Rübsamen-Waigmann H, Strebhardt K: **Induction and down-regulation of PLK, a human serine/threonine kinase expressed in proliferating cells and tumors.** *Proc Natl Acad Sci USA* 1994, **91**:1736-1740.
57. Wang H, Gauthier BR, Hagenfeldt-Johansson KA, Iezzi M, Wollheim CB: **Foxa2 (HNF3beta) controls multiple genes implicated in metabolism-secretion coupling of glucose-induced insulin release.** *J Biol Chem* 2002, **277**:17564-17570.
58. Mincheva A, Lichter P, Schütz G, Kaestner KH: **Assignment of the human genes for hepatocyte nuclear factor 3-alpha, -beta, and -gamma (HNF3A, HNF3B, HNF3G) to 14q12-q13, 20p11, and 19q13.2-q13.4.** *Genomics* 1997, **39**:417-419.
59. Wolfrum C, Shih DQ, Kuwajima S, Norris AW, Kahn CR, Stoffel M: **Role of Foxa-2 in adipocyte metabolism and differentiation.** *J Clin Invest* 2003, **112**:345-356.
60. Kim TM, Yim SH, Shin SH, Xu HD, Jung YC, Park CK, Choi JY, Park WS, Kwon MS, Fiegler H, Carter NP, Rhyu MG, Chung YJ: **Clinical implication of recurrent copy number alterations in hepatocellular carcinoma and putative oncogenes in recurrent gains on 1q.** *Int J Cancer J Int Cancer* 2008, **123**:2808-2815.
61. Kauffmann A, Rosselli F, Lazar V, Winnepeninckx V, Mansuet-Lupo A, Dessen P, van den Oord JJ, Spatz A, Sarasin A: **High expression of DNA repair pathways is associated with metastasis in melanoma patients.** *Oncogene* 2008, **27**:565-573.
62. McCaughan GW, Gorrell MD, Bishop GA, Abbott CA, Shackel NA, McGuinness PH, Levy MT, Sharland AF, Bowen DG, Yu D, Slaitini L, Church WB, Napoli J: **Molecular pathogenesis of liver disease: an approach to hepatic inflammation, cirrhosis and liver transplant tolerance.** *Immunol Rev* 2000, **174**:172-191.
63. Enk CD, Jacob-Hirsch J, Gal H, Verbovetski I, Amariglio N, Mevorach D, Ingber A, Givol D, Rechavi G, Hochberg M: **The UVB-induced gene expression profile of human epidermis in vivo is different from that of cultured keratinocytes.** *Oncogene* 2006, **25**:2601-2614.
64. Wei RR, Huang GL, Zhang MY, Li BK, Zhang HZ, Shi M, Chen XQ, Huang L, Zhou QM, Jia WHJ, Zheng XFS, Yuan YF, Wang HY: **Clinical significance and prognostic value of microRNA expression signatures in hepatocellular carcinoma.** *Clin Cancer Res Off J Am Assoc Cancer Res* 2013.
65. Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JGM, Foekens JA, Martens JWM: **Subtypes of breast cancer show preferential site of relapse.** *Cancer Res* 2008, **68**:3108-3114.
66. Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, Cui Q: **An Analysis of Human MicroRNA and Disease Associations.** *PLoS ONE* 2008, **3**:e3420.
67. Zeng L, Yu J, Huang T, Jia H, Dong Q, He F, Yuan W, Qin L, Li Y, Xie L: **Differential combinatorial regulatory network analysis related to venous metastasis of hepatocellular carcinoma.** *BMC Genomics* 2012, **13** Suppl 8: S14.
68. Tu K, Yu H, Hua YJ, Li YY, Liu L, Xie L, Li YX: **Combinatorial network of primary and secondary microRNA-driven regulatory mechanisms.** *Nucleic Acids Res* 2009, **37**:5969-5980.
69. Chen M, Xiao J, Zhang Z, Liu J, Wu J, Yu J: **Identification of human HK genes and gene expression regulation study in cancer from transcriptomics data analysis.** *PLoS One* 2013, **8**:e54082.

doi:10.1186/1471-2164-15-S1-S13

Cite this article as: Wei et al.: Application of microRNA and mRNA expression profiling on prognostic biomarker discovery for hepatocellular carcinoma. *BMC Genomics* 2014 **15**(Suppl 1):S13.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

