

## Peripheral blood mRNA expression patterns to differentiate hepatocellular carcinoma from other hepatic diseases

PJ Zhang<sup>1</sup>, W Run<sup>1</sup>, Liang P<sup>2</sup>, CB Wang<sup>3</sup>, XX Deng<sup>1</sup>, B Wang<sup>1</sup>, B Chen<sup>1</sup>, Jiao J<sup>1</sup>, HY Liu<sup>1</sup>, ZN Dong<sup>1</sup>, XJ Zhang<sup>1</sup>, YP Tian<sup>1</sup>

<sup>1</sup>Department of Clinical Biochemistry, Chinese PLA General Hospital, Beijing, China, <sup>2</sup>Department of Interventional Ultrasound, Chinese PLA General Hospital, Beijing, China, <sup>3</sup>Department of Clinical Laboratory Medicine, Chinese PLA General Hospital, Beijing, China

### TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Material and methods
  - 3.1. Patients
  - 3.2. Blood collection and RNA isolation
  - 3.3. Genes and GeXP primer design
  - 3.4. GeXP multiplex RT-PCR
  - 3.5. GeXP multiplex data analysis
  - 3.6. Diagnosis model for hepatocellular carcinoma development
4. Results
  - 4.1. RT-PCR of single gene by chimeric and universal primers
  - 4.2. Multiplex primer RT-PCR and optimization
  - 4.3. Gene expressions between five different groups
  - 4.4. Diagnosis model to differentiate the five groups
  - 4.5. Workflow for the diagnosis model by GeXP
5. Discussion
6. Acknowledgments
7. References

## 1. ABSTRACT

Peripheral blood genes expressions profiling (GeXP) have been convinced to be more specific for the diagnosis of cancer and other diseases, and the GeXP system provides an ideal method to analyze multiple genes expression in one normalized and equable system. We aim to differentiate hepatocellular carcinoma from other hepatic diseases based on peripheral blood and the GeXP system. Fifteen selected hepatic diseases related genes with two house-keeping genes for normalization were detected by the GeXP system. The diagnosis model was based on K nearest neighbor classifier and cross validation, and software based on MATLAB software was built for differential diagnosis of hepatic diseases. Eight hepatic related genes were demonstrated to show an obvious statistic difference in expressions while the K nearest neighbors classifier showed that the accuracy for normal controls, hepatitis B, liver cirrhosis, hepatocellular carcinoma and the Other group was separately 80.57%, 78.17%, 84.48%, 73.24% and 85.85%. The set of validation has been carried out to assess the accuracy of Model Two and the accuracy was even higher than the set of building for the model, except for the hepatitis B (HBV) group. A sensitive and specific GeXP system of eight genes has been developed for the accurate differential diagnosis of hepatic disease.

## 2. INTRODUCTION

Hepatocellular carcinoma (HCC) is one of the most common cancers in the world with over half a million deaths annually (1). It shows a significant relation with the prevalence of risk factors, such as chronic infection with hepatitis B or C virus (HBV, HCV), liver cirrhosis (LC) and alcohol abuse. Chronic HBV and HCV are attributed to three-quarters of HCC development worldwide and especially in China, > 90% of HCC cases are reported to have HBV background (2). It is widely recognized chronic carriers of HBV is the leading cause of LC and HCC (3). HCC is often diagnosed at an advanced stage, so there is no effective treatment available for these patients. Early diagnosis can provide a better chance for treatment, however, even if AFP, a well known biomarker for HCC, had only a lower sensitivity and specificity for the presence of HCC. In addition, ultrasound (US), computed tomography (CT) and magnetic resonance imaging (MRI) are limited by their accuracy (4). The need for improved methods to accurately differentiate HCC from other hepatic diseases is desirable.

Many studies have been reported to use peripheral blood cells transcriptome analysis for disease prediction and cancer classification (5-7). In addition, peripheral blood are easily obtained and considered to be

**Table 1.** Clinical characteristics of all the peripheral blood samples

Clinical characteristics	NC (54)	HBV (42)	LC (24)	HCC (103)	Others (64)
Age (years)	58	50.5	52	56	41
AFP (ng/ml)	1.74	2.21	7.61	133.40	2.66
ALB (g/L)	47.40	40.80	35.80	41.80	42.30
ALT (U/L)	20.20	38.17	41.70	41.50	25.50
AST (U/L)	20.50	44.32	54.90	39.00	20.95
GGT (U/L)	27.35	21.25	88.30	59.45	34.30
ALP (U/L)	67.40	58.35	148.80	84.80	68.80

The clinical characteristics were presented as median values

minimally invasive with a standard operation procedure. Both of them make blood samples an attractive alternative for the HCC development diagnostic procedures. Transcriptome analysis surveys thousands of genes expression in peripheral blood cells, however, only 20-100 genes have discriminatory power (8). HCC is a complex multigene or multifactorial disease, it is obvious that a single biomarker will be not adequate to reflect the status of HCC and a panel of biomarkers is needed for differential diagnosis of hepatic disease (9).

The Beckman Coulter (Fullerton, CA, USA) GenomeLab GeXP Genetic Analysis system, known as GeXP, is designed ideally to run up to 35 genes per reaction containing 5 ng to 500 ng of total RNA and can be used to interrogate 192 samples simultaneously in a single analysis (10). The GeXP system uses a combined gene-specific, universal priming strategy that converts multiplexed PCR to a two-primer process using universal primers. This strategy overcomes the variations in amplification efficiency of different genes. A capillary electrophoresis separation system, based on the products size, ensures it is sensitive enough to precisely detect even small changes in gene expression. The mechanism of the GeXP system was observed (Figure 1). Based on the strength of the GeXP system, it can serve as a sensitive and precise panel of genes expression analysis, which may serve as diagnostic models for different diseases.

In our study, the GeXP analysis system is used to detect a panel of genes expression in peripheral blood samples of HCC and different hepatic diseases. We aim to set up a blood-based, disease-specific diagnostic screening method to accurately differentiate HCC from other hepatic diseases.

### 3. MATERIALS AND METHODS

#### 3.1. Patients

With patient consent, peripheral blood samples of 54 healthy normal controls (NC), 42 Chronic HBV-infected patients (HBV), 24 patients with liver cirrhosis (LC), 103 HCC patients with HBV or cirrhosis and 64 other liver disease patients (Others) were collected from our hospital. Clinical characteristics of all the peripheral blood samples were observed (Table 1). The five groups were randomly split into a set of building for model and a set of validation. The set of building for model included 36 NC controls, 28 HBV, 16 LC, 81 HCC and 44 Others patients. The set of validation included 18 NC controls, 15 HBV, 8 LC, 27 HCC and 21 Others patients. Peripheral blood samples from HBV and HCC patients were collected before any

treatment such as surgery, chemotherapy and radiation therapy. Peripheral blood samples were aliquoted and stored at -20°C.

#### 3.2. Blood collection and RNA isolation

2.5 mL peripheral blood was collected from each sample and directly poured into PAXGene blood RNA tubes (Qiagen, Valencia, CA), which contained 6.9 ml additive to stabilize mRNA in whole blood. After blood collection, the PAXGene blood RNA tubes were gently inverted several times and were stored at -20 °C for 24 hours and then were transferred to -80 °C until RNA extraction. Total RNA was isolated according to the manufacturer instructions using the PAXGene Blood RNA kit (Qiagen, Valencia, CA). The quality of RNA was quantified by UV absorbance measurement using a DU 800 spectrophotometer (Beckman Coulter, Fullerton, CA). The quality of RNA was determined visually by inspecting the integrity of 28S and 18S ribosomal bands after agarose gel electrophoresis.

#### 3.3. Genes and GeXP primer design

15 hepatic diseases related genes (GPC3, Glypican-3; HGF, hepatocyte growth factor; RPS24, 40S ribosomal protein S24; ANXA1, annexin1; FOS, v-fos FBJ murine osteosarcoma viral oncogene homolog; SPAG9, sperm associated antigen 9; HSPA1B, heat shock protein A1B; PFDN5, Prefoldin subunit 5; IL8, interleukin-8; GOS2, G<sub>0</sub>/G<sub>1</sub>switch 2; CXCR4, chemokine receptor 4; RPL27, ribosomal protein L27; PFN1, profilin-1; CALR, calreticulin; GZMA, Granzyme A) and 2 housekeeping genes (ACTB, actin, beta; B2M, beta 2 microglobulin) were chosen for building the panel of differentiating HCC development. The genes and primers were identified (Table 2). Primers were designed by GenomeLab GeXP eXpress Profiler software. Primers were also designed for Kanamycin RNA intern positive control with 325 bp PCR product. In the later cycles of PCR, fluorescently labeled universal forward primer was used. All primers were obtained from SBS Genetech (Beijing, China) and stocked with concentration 100 µM. Primers were designed to yield PCR product ranging from 117 to 284 bp.

#### 3.4. GeXP multiplex RT-PCR

In our multiplex RT-PCR experiment, 50 ng of RNA was used as starting material for reverse transcription (RT) with chimeric reverse primers in single reaction. The RT reactions were performed according to the manual of the Genome Lab GeXP Start Kit. The concentration of adjustment primers varied from different dilution rate (1:2, 1:4, 1:8, 1:16, 1:32, 1:64, 1:128, 1:256). RT reactions were incubated at 48 °C for 1 minute, 42 °C for 60 minutes, 95

A. Reverse gene-specific primer with universal sequences (non-labeled)



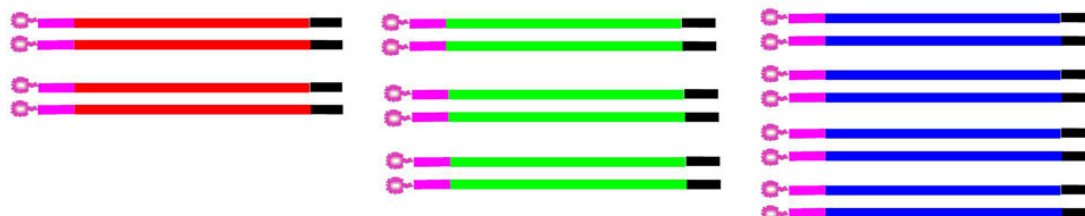
B. Forward gene-specific primer with universal sequences (non-labeled)



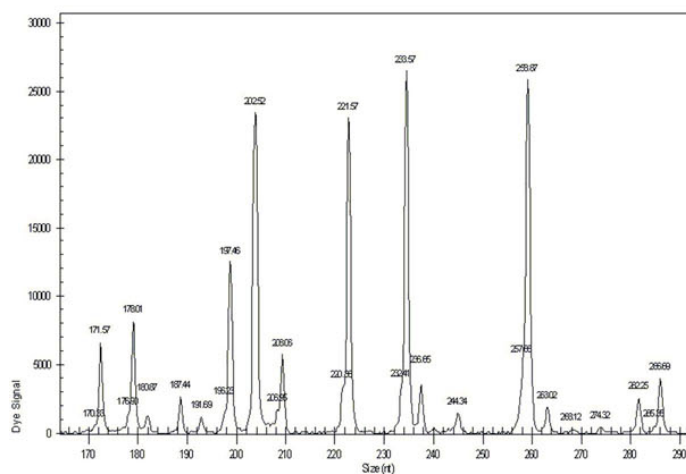
C. Multiplex PCR with labeled universal forward sequences



D. Multiplex PCR product



E. Multiplex PCR product analyzed by capillary electrophoresis



**Figure 1.** Thymosin-alpha1 elevated the survival rate of sepsis mice induced by CLP. The survival rate of THI mice was significantly higher than that of CLP group after 24 hours, (\*) indicates a  $P$  value less than 0.05 ( $\chi^2 = 4.96$ ,  $P = 0.026$  at 24 h;  $\chi^2 = 4.59$ ,  $P = 0.032$  at 48 h;  $\chi^2 = 4.53$ ,  $P = 0.033$  at 72 h).

**Table 2.** Genes and primers used in the GeXP

Gene symbol	RefSeq	Reverse primer	Forward primer	Size of product
GPC3	L47125	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	117
HGF	M77227	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	162
RPS24	BC000523	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	172
ANXA1	NM_000700	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	177
FOS	BC004490	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	182
SPAG9	AK023512	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	187
HSPA1B	NM_005346	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	192
PFDN5	NM_002624	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	197
ACTB <sup>1</sup>	NM_001101	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	202
IL8	NM_000584	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	207
GOS2	NM_015714	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	222
CXCR4	AF348491	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	232
RPL27	NM_000988	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	237
PFN1	NM_005022	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	257
CALR	NM_004343	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	262
GZMA	NM_006144	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	279
B2M <sup>1</sup>	NM_004048	GTACGACTCACTATAGGGAATCATTCATCACCAGAGCC	AGGTGACACTATAGAATACTGGATGAGGAAGGGTTTGA	284

Underlined oligonucleotides are universal sequences. <sup>1</sup> Gene used for normalization.

°C for 5 minutes and held at 4 °C. PCR reaction was then performed in a Thermo-Fast 96-well PCR Detection Plate containing fluorescently labeled universal primer and unlabeled reverse universal primer. PCR reaction was performed under the conditions: 95 °C for 10 minutes, followed by 35 cycles of 94 °C for 30 seconds, 55 °C for 30 seconds and 68 °C for 1 minute.

### 3.5. GeXP multiplex data analysis

PCR products from multiplex primers were prepared for capillary electrophoresis. 2 µl of PCR product was added to the 96-well detection plate containing 37.5 µl of sample loading solution along with 0.5 µl of DNA Size Standard 400. The GeXP Genetic Analysis System matched each fragment size with the specific gene and measured the fluorescent dye signal strength in arbitrary units (A.U.). Then the data was exported from the GeXP Genetic Analysis System and normalized to the housekeeping gene B2M. The relative genes expression data were normalized by dividing the area under the curve (AUC) by the AUC of B2M; the data were then log transformed. GeXP results among the five groups were compared through the use of nonparametric statistics (Mann-Whitney U tests).

### 3.6. Diagnosis model for hepatocellular carcinoma development

The missing genes expression detected by GeXP were imputed with k-nearest neighbors imputation using k=3. Differentially expressed genes were identified by statistically significant differences ( $p < 0.05$ ) between the five groups using one-way ANOVA analysis and were used for building the diagnosis model for HCC development. The diagnosis model was based on K nearest neighbors classifier (KNN) and 5-fold cross validation (4/5) of the samples as the training set and the remainder 1/5 of the samples as the testing set (11). The cross validation with iteration was repeated at a number of 100 times and the average accuracy for the five groups was calculated, then the optimized model was chosen for building the diagnosis model and the accuracy of the validation set was calculated. In the final step, software based on MATLAB software was developed to predict the diagnosis model for HCC development.

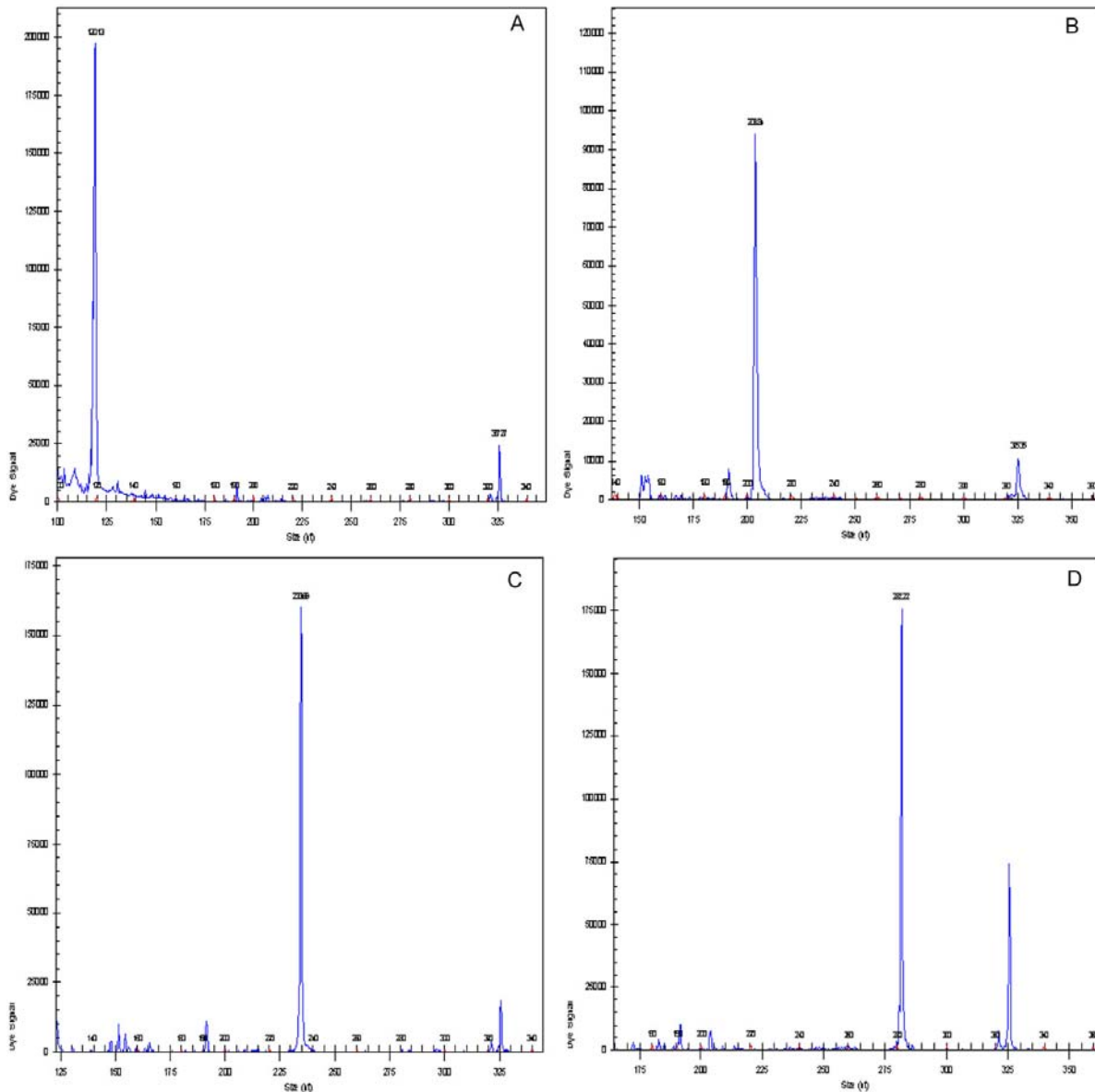
## 4. RESULTS

### 4.1. RT-PCR of single gene by chimeric and universal primers

At the first stage of multiplex RT-PCR, each gene-specific PCR product was analyzed by GeXP to verify the expected peak fragment and assess any unintended PCR product. The initial data analyzed by the GeXP system from the genes-specific primer was observed (Figure 2). The seventeen genes fragment size detected by GeXP all matched the actual size designed for PCR product. GPC3, HGF, RPS24, ANXA1, FOS, SPAG9, HSPA1B, PFDN5, ACTB, IL8, GOS2, CXCR4, RPL27, PFN1, CALR, GZMA and B2M gene was approximately 117 bp, 162 bp, 172 bp, 177 bp, 182 bp, 187 bp, 192 bp, 197 bp, 202 bp, 207 bp, 222 bp, 232 bp, 237 bp, 257 bp, 262 bp, 279 bp and 284 bp. The Kanamycin RNA intern positive control with 325 bp was all detected in the seventeen single gene-specific primer. The nonspecific genes were not detected, indicating an absence of contaminating DNA.

### 4.2. Multiplex primer RT-PCR and optimization

The 17 genes were divided into two groups: adjustment group and non-adjustment group. The adjustment group included ACTB and B2M genes, the other genes were the non-adjustment group. The dilution rate of ACTB and B2M genes were separately 1:2, 1:4, 1:8, 1:16, 1:32, 1:64, 1:128 and 1:256. The initial data analyzed by the GeXP system from multiplex primer of 1:2, 1:8, 1:32, 1:128 were observed (Figure 3), 1:4, 1:16, 1:64 and 1:256 were not observed. In the dilution rate of 1:2, the gene of ACTB and B2M showed a higher dye signal than the other genes, the two genes may have no sufficient potential dynamic range. In the dilution rate of 1:4 and 1:256, some genes such as ? showed the lower gene signal was not detected by the GeXP system. In the dilution rate of 1:8, all of the seventeen genes were detected. The lowest gene signal could be detected and the highest gene signal had sufficient potential dynamic range. In the dilution rate of 1:16 and 1:128, the genes showed desired levels were less than 1:8. In the dilution rate of 1:32 and 1:64, the background noise level was so high that it may have affected the results detected by the GeXP system. Compared to the eight dilution rates, 1:8 was the most appropriate dilution rate for the multiplex primer RT-PCR.

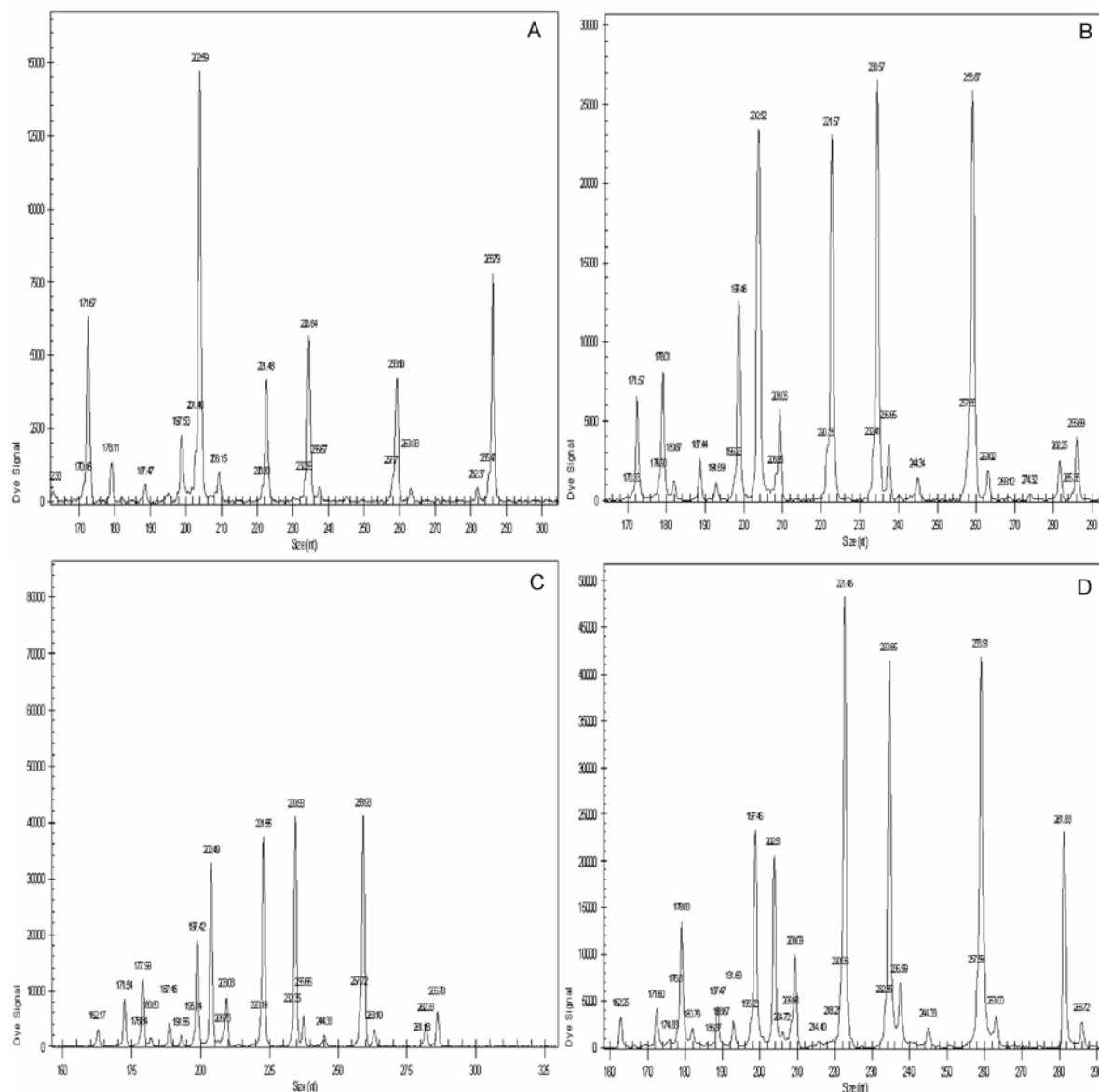


**Figure 2.** RT-PCR of single gene by chimeric and universal primers. RT-PCR of each gene-specific PCR product was detected by GeXP to analyze the specificity of 17 genes. A Kanamycin RNA with 325 bp served as an intern positive control. (A), GPC3 with 117 bp. (B), ACTB with 202 bp. (C), CXCR4 with 232 bp. (D), GZMA with 279 bp. The other genes were not shown. All of the genes fragment size matched the actual size designed for the PCR product and the intern positive control was completely detected.

#### 4.3. Genes expressions between five different groups

The GZMA and IL8 genes were missing in most of the samples, so the two genes were removed from the panel of genes expressions. After one-way ANOVA analysis, the nine genes of CALR, PFN1, SPAG9, ANXA1, HGF, FOS, GPC3, HPSA1B and CXCR4 showed significant differences ( $p < 0.01$ ) between the five groups (Figure 4). The four genes of RPS24, RPL27, PFDN5 and GOS2 showed no significant differences ( $p < 0.01$ ) between the five groups, so the four genes were removed from the panel of genes expressions. For CLAR gene, the NC, HBV

and LC group showed no significant difference when compared to each other. For PFN1 gene, the HBV group showed no significant difference when compared to the LC and HCC group. For SPAG9 gene, it cannot differentiate the HBV, LC and HCC group. For ANXA1 gene, the NC, HBV, LC and Other group showed no significant difference when compared to each other. For HGF gene, compared to LC group, the other four groups showed a significant difference, while the other four groups showed no significant difference when compared to each other. For FOS gene, it cannot differentiate the NC, HBV and HCC



**Figure 3.** Multiplex primer RT-PCR and optimization. The 17 genes were divided into an adjustment group and non-adjustment group. The adjustment group was ACTB and B2M genes, (A-D) showed a different dilution rate. (A), 1:2 showed a higher dye signal, so it may have no sufficient potential dynamic range. (B), 1:8, all of the 17 genes were detected. The lowest gene signal could be detected and the highest gene signal had sufficient potential dynamic range. (C) 1:32, the background noise level was higher, which could affect the genes expression. (D) 1:128, the genes showed desired levels were less than 1:8. Compared to the eight dilution rates, 1:8 was the most appropriate dilution rate for multiplex primer RT-PCR.

group. For GPC3 gene, it cannot classify the HBV, LC and HCC group. For HSPA1B gene, the NC, HBV and Other group showed no significant difference. For CXCR4 gene, it cannot classify the NC, HBV, LC and HCC groups at all. The nine single gene showed significant difference between the five groups, however, it cannot differentiate the five groups accurately.

#### 4.4. Diagnosis model to differentiate the five groups

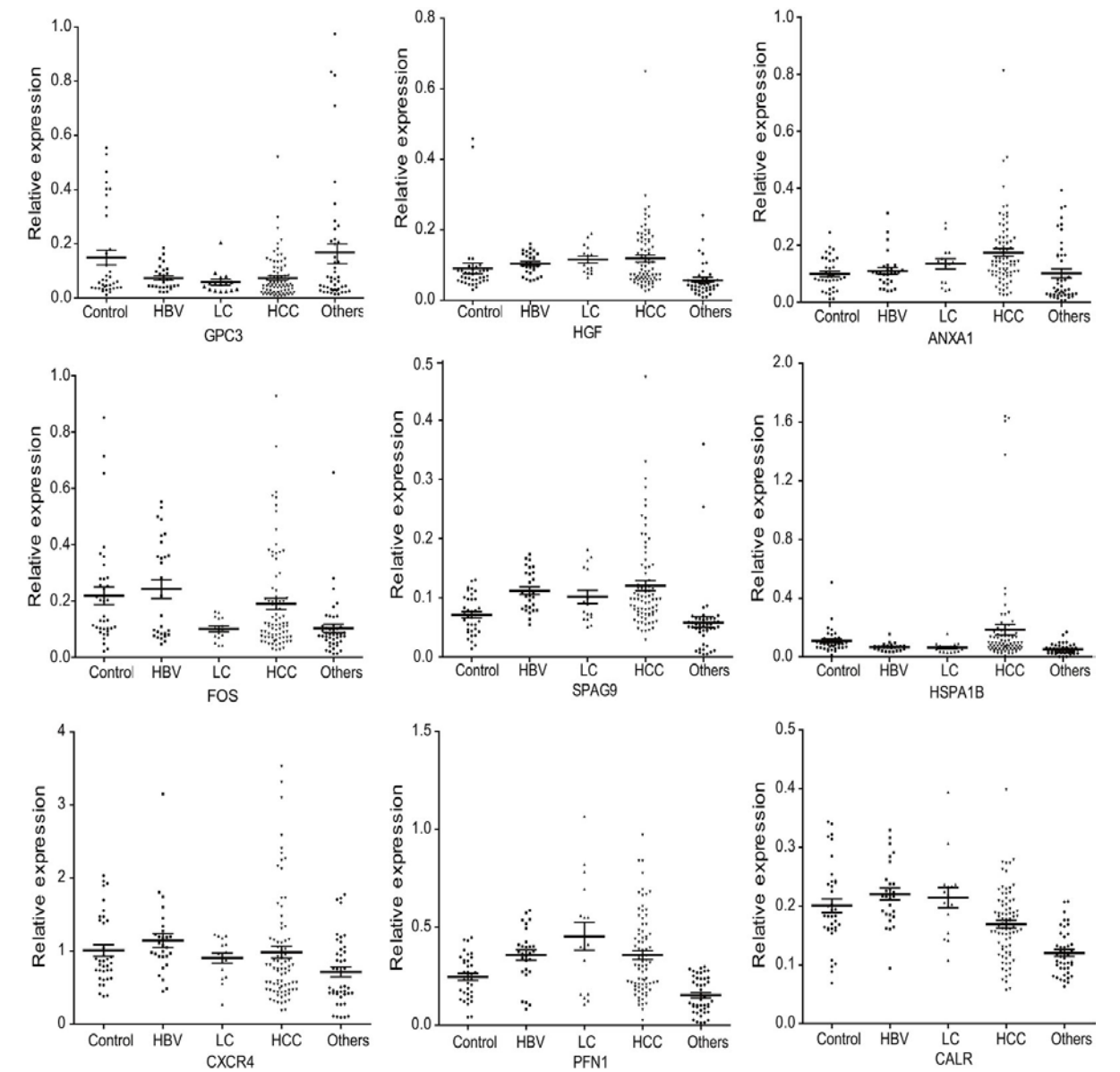
The nine genes were arranged ascending according to the P value, and then five models were built to

compare the accuracy of differentiating the five groups. Model one was a nine genes profiling of CALR, PFN1, SPAG9, ANXA1, HGF, FOS, GPC3, HSPA1B and CXCR4 gene. Model two was an eight genes profiling of CALR, PFN1, SPAG9, ANXA1, HGF, FOS, GPC3 and HSPA1B gene. Model three was a seven genes profiling of CALR, PFN1, SPAG9, ANXA1, HGF, FOS and GPC3 gene. Model four was a six genes profiling of CALR, PFN1, SPAG9, ANXA1, HGF and FOS gene. Model five was a five genes profiling of CALR, PFN1, SPAG9, ANXA1 and HGF gene. The average accuracy of the five

**Table 3.** The average accuracy of classifying the five groups by KNN and cross validation (%)

Model	Genes profiling	NC	HBV	LC	HCC	Others
Model One	CALR,PFN1,SPAG9,ANXA1,HGF,FOS,GPC3,HPSA1B,CXCR4	70.13	70.77	83.82	66.55	81.52
Model Two	CALR,PFN1,SPAG9,ANXA1,HGF,FOS,GPC3,HPSA1B	80.57	78.17	84.48	73.24	85.85
Model Three	CALR,PFN1,SPAG9,ANXA1,HGF,FOS,GPC3	85.52	80.95	84.35	71.75	84.00
Model Four	CALR,PFN1,SPAG9,ANXA1,HGF,FOS	81.48	72.70	83.85	73.18	85.52
Model Five	CALR,PFN1,SPAG9,ANXA1,HGF	77.90	74.47	79.05	70.38	88.50

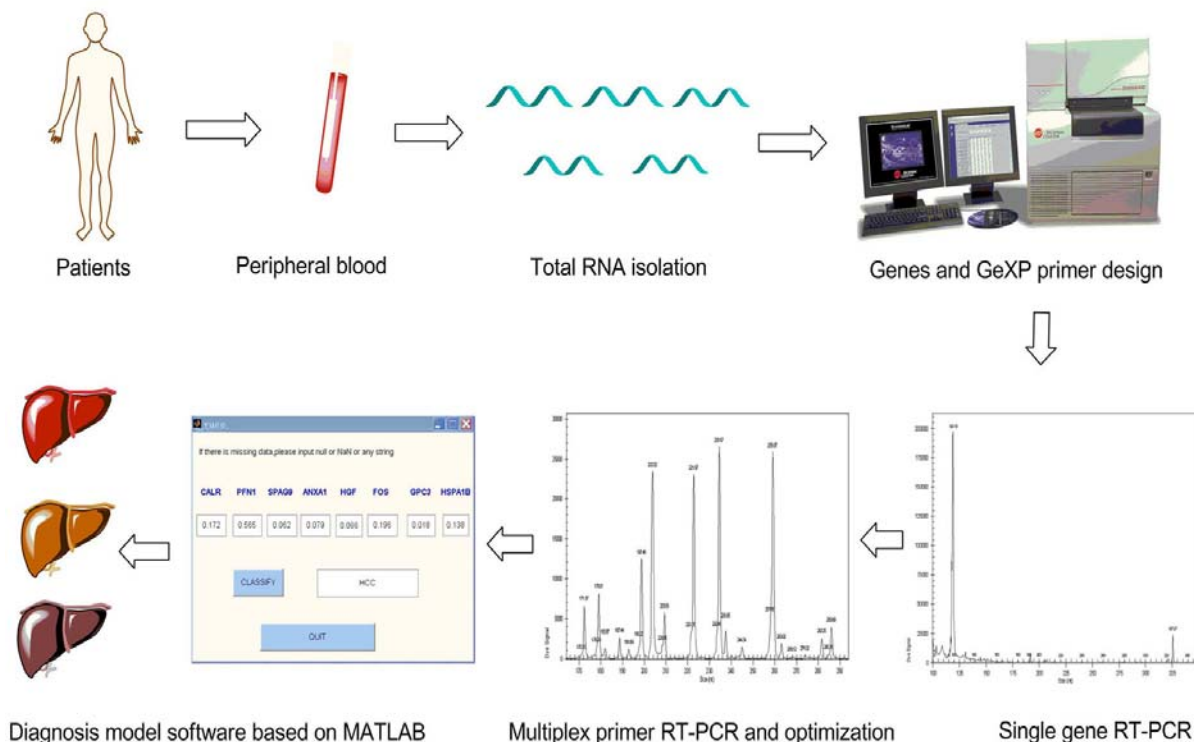
Five models were built to compare the accuracy of differentiating the NC, HBV, LC, HCC and Others groups. The average accuracy of classifying the NC, HBV, LC, HCC and Other group by Model one were 70.13%, 70.77%, 83.82%, 66.55% and 81.52%. Model two were 80.57%, 78.17%, 84.48%, 73.24% and 85.85%. Model three were 85.52%, 80.95%, 84.35%, 71.75% and 84.00%. Model four was 81.48%, 72.70%, 83.85%, 73.18% and 85.52%. Model five was 77.90%, 74.47%, 79.05%, 70.38% and 88.50%. After comparing the five models, Model two had the best accuracy of differentiating the five groups.



**Figure 4.** The genes expression showed a significant difference between the five groups detected by the GeXP System. After one-way ANOVA analysis, the nine genes of CALR, PFN1, SPAG9, ANXA1, HGF, FOS, GPC3, HPSA1B and CXCR4 showed significant differences ( $p<0.01$ ) between the NC, HBV, LC, HCC and Others groups. The nine single gene showed a significant difference between the five groups, however, it cannot differentiate the five groups accurately.

models was recorded (Table 3). The average accuracy of classifying the NC, HBV, LC, HCC and Other group by Model one were 70.13%, 70.77%, 83.82%, 66.55% and 81.52%. Model two were 80.57%, 78.17%, 84.48%, 73.24% and 85.85%. Model three were 85.52%, 80.95%, 84.35%, 71.75% and 84.00%. Model four was 81.48%, 72.70%, 83.85%, 73.18% and 85.52%. Model five was 77.90%, 74.47%, 79.05%, 70.38% and 88.50%.

## Hepatocellular carcinoma diagnosis



**Figure 5.** A workflow for diseases or cancer diagnosis model based on the peripheral blood and GeXP system. Total RNA of peripheral blood was isolated by the PAXgene Blood RNA kit. Genes related to the disease or cancer was chosen and primers were designed by GeXP. The GeXP detection system involved several steps. Firstly, single gene RT-PCR was detected to analyze the specificity. Secondly, the Multiplex primer RT-PCR and optimization were done to detect genes expression. Thirdly, after one-way ANOVA analysis, we got the differentially expressed genes. Finally, a diagnosis model to differentiate diseases or cancer was built based on KNN and cross validation. Lastly, the model was chosen for building the diagnosis model software.

73.24% and 85.85%. Model three were 85.52%, 80.95%, 84.35%, 71.75% and 84.00%. Model four was 81.48%, 72.70%, 83.85%, 73.18% and 85.52%. Model five was 77.90%, 74.47%, 79.05%, 70.38% and 88.50%. After comparing the five models, Model two had the best accuracy of differentiating the five groups, and the set of validation was used to assess the accuracy of Model two by the software we built. The average accuracy of differentiating the five groups was separately 83.33%, 73.33%, 100%, 75.00% and 95.24%. The accuracy of the set of validation was higher than the set of building for the model except for the HBV group.

### 4.5. Workflow for the diagnosis model by GeXP

A workflow for diseases or cancer diagnosis model was built based on the GeXP system (Figure 5). Genes related to the disease or cancer was chosen to analyze by GeXP. Firstly, RT-PCR of single gene was detected to analyze the specificity. Secondly, the Multiplex primer RT-PCR was optimized individually by attenuation to the most appropriate dilution rate. Thirdly, the genes were compared by one-way ANOVA analysis to get the differentially expressed genes. Finally, KNN and cross validation were used to build a diagnosis model to differentiate diseases or the cancer. Lastly, the model was chosen for building the diagnosis model software.

## 5. DISCUSSION

Because peripheral blood was widely used, easily obtained and minimally invasive, it was believed to be an attractive type for clinical research. In addition, mRNA in peripheral blood samples was demonstrated to be stable in whole peripheral blood samples (12). Specific mRNA concentrations in EDTA blood stored tubes changed, and it could affect the results of PCR based techniques. These changes were eliminated or markedly reduced when stored in PAX gene tubes, which was recommended (13). Studies had shown that peripheral blood had the potential of early detection (14-16). In our study, based on the peripheral blood and GeXP system, we aimed to build a simple, rapid, efficient and standardized detection method.

The GeXP system had distinct advantages; however, there were also some limitations. The number of genes analyzed was limited to 35, and the gene size arranged from 150 to 350 bp. In addition, the dynamics of single gene RT-PCR reactions differed from those in multiplexed reactions, multiplex RT-PCR had to be optimized by dilution. There are several steps involved in the procedure of the GeXP system; changes in any step may improve the results of the GeXP system. A standardized detection method was built in our study to



analyze genes profiling in HCC development. GeXP was demonstrated to be a sensitive and effective assay for infection with the pandemic influenza A H1N1 virus and was able to detect seasonal influenza H3N2 coinfection (17). 68 unique varicella zoster virus gene transcripts in five GeXP multiplex RT-PCR was also detected to be rapid and sensitive (18). The GeXP system was also used to detect gene profiling expressions (19-21).

CALR acted as a major Calcium-binding protein of the endoplasmic reticulum and transcriptional regulator. It was overexpressed in adenocarcinoma cells (22); it was also correlated with intracellular iron increase and protection from oxidative stress (23). In our study, there was no significant difference between the NC group and HCC group. There were so few research studies on the relationship between CALR and HCC; it was unclear how CALR took part in the progression and its correlation with degree of liver injury. PFN1 was regarded as a tumor-suppressor molecule for breast cancer by enhancing ADP-to-ATP exchange on G-actin. Studies had demonstrated that PFN1 was over-expression in cancer cells by upregulating PTEN and suppressing AKT activation, and it was also a negative regulator of mammary carcinoma aggressiveness (24). If the PFN1 was silenced, it could inhibit endothelial cell migration, proliferation, and cord morphogenesis (25). We found the PFN1 gene expression in the HCC group was higher than the normal group; however, it was lower than the LC group. It implied that PFN1 may involve the mechanism of HCC. SPAG9 was a scaffolding protein that connected protein kinases to their transcription factor targets together for the activation of specific signaling pathways, such as a JNK signal (26). It can act as a biomarker for breast cancer and cervical carcinoma (27, 28). We demonstrated that SPAG9 in the HCC group was higher than the NC group. It may exist as a potential biomarker for HCC. ANXA1 was a member of the annexin family of phospholipids binding and calcium-binding proteins (29). It was overexpressed in breast cancer and HCC (30, 31). This was mainly because ANXA 1 was involved in tumorigenesis, and our study demonstrated it was higher in the HCC group than the NC group. HGF was found to stimulate growth of melanocytes and endothelial cells, however, it inhibited the growth of hepatocellular carcinoma cells (32). Sera of patients with acute hepatitis, chronic hepatitis and liver cirrhosis contained elevated levels of HGF (33). This was because persistent liver injury was closely related to the occurrence of HCC. Fos was an oncogene which was upregulated in response to many extracellular signals, and HBV infection also affected the expression of Fos (34). We found the NC group showed no difference with HCC group mainly due to the fact the Fos can be affected by many signal pathways. GPC3 was highly expressed in a hepatoma cell, and it was proposed as a potential novel tumor marker for HCC (35), however, in our study the mRNA expression of GPC3 in peripheral blood was lower than the NC group. This may be because the peripheral blood was different from the tissue, and many more samples would be needed to validate the results. HSPA1B was a gene that interacted with other heat shock proteins to stabilize existing proteins against aggregation and mediated the translated proteins (36). It was also involved in the

rapid formation of a transcription-competent state during the minor ZGA through interaction with the RNA polymerase II. HSPA1B in the HCC group was higher than the LC or HBV group; however, the mechanism was not well understood. CXCR4 was a gene described to regulate the homing of lymphocytes in inflammatory tissues. High CXCR4 expression was related to tumor dissemination in colorectal, breast and hepatocellular carcinoma (37-39). In our study, CXCR4 showed no difference between NC, HBV, LC and the HCC group, however, the Other group showed a significant difference when compared to the other four groups. It was mainly due to the difference between peripheral blood and tissue, and it could also be explained by the different mechanism of different liver diseases.

In conclusion, peripheral blood samples were chosen as the diagnosis type because it was easily obtained, minimally invasive and performed with standard operation procedure. GeXP was chosen as the detection system because it was sensitive, precise and convenient for gene expression analysis. The KNN and cross validation were used to build a diagnosis model to classify a gene as a disease or cancer. An eight genes panel of CALR, PFN1, SPAG9, ANXA1, HGF, FOS, GPC3 and HSPA1B was built for differentiating NC, HBV, LC, HCC and Others group and the accuracy was separately 80.57%, 78.17%, 84.48%, 73.24% and 85.85%.

## 6. ACKNOWLEDGMENTS

The first 2 authors contributed equally to this article. This study was supported by the Ministry of Science and Technology of China (2006FY230300).

## 7. REFERENCES

1. D. M. Parkin, F. I. Bray and S. S. Devesa: Cancer burden in the year 2000. The global picture. *Eur J Cancer*, 37 Suppl 8, S4-66 (2001)
2. J. D. Gao, Y. F. Shao, Y. Xu, L. H. Ming, Z. Y. Wu, G. T. Liu, X. H. Wang, W. H. Gao, Y. T. Sun, X. L. Feng, L. M. Liang, Y. H. Zhang and Z. T. Sun: Tight association of hepatocellular carcinoma with HBV infection in North China. *Hepatobiliary Pancreat Dis Int*, 4, 46-49 (2005)
3. I. Merican, R. Guan, D. Amarapuka, M. J. Alexander, A. Chutaputti, R. N. Chien, S. S. Hasnain, N. Leung, L. Lesmana, P. H. Phiet, H. M. Sjalfoellah Noer, J. Sollano, H. S. Sun and D. Z. Xu: Chronic hepatitis B virus infection in Asian countries. *J Gastroenterol Hepatol*, 15, 1356-1361 (2000)
4. S. A. K. Asmaa I Gomaa, Edward LS Leen, Imam Waked, Simon D Taylor-Robinson: Diagnosis of hepatocellular carcinoma. *World J Gastroenterol*, 15, 1301-1314 (2009)
5. P. R. Sinnaeve, M. P. Donahue, P. Grass, D. Seo, J. Vonderscher, S. D. Chibout, W. E. Kraus, M. Sketch, Jr., C. Nelson, G. S. Ginsburg, P. J. Goldschmidt-Clermont and C. B. Granger: Gene expression patterns in peripheral

- blood correlate with the extent of coronary artery disease. *PLoS One*, 4, e7037 (2009)
6. B. Mesko, S. Poliska, A. Szegedi, Z. Szekanecz, K. Palatka, M. Papp and L. Nagy: Peripheral blood gene expression patterns discriminate among chronic inflammatory diseases and healthy controls and identify novel targets. *BMC Med Genomics*, 3, 15 (2010)
7. M. E. Burczynski, N. C. Twine, G. Dukart, B. Marshall, M. Hidalgo, W. M. Stadler, T. Logan, J. Dutcher, G. Hudes, W. L. Trepicchio, A. Strahs, F. Immermann, D. K. Slonim and A. J. Dorner: Transcriptional profiles in peripheral blood mononuclear cells prognostic of clinical outcomes in patients with advanced renal cell carcinoma. *Clin Cancer Res*, 11, 1181-1189 (2005)
8. D. Singh, P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, C. Ladd, P. Tamayo, A. A. Renshaw, A. V. D'Amico, J. P. Richie, E. S. Lander, M. Loda, P. W. Kantoff, T. R. Golub and W. R. Sellers: Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, 1, 203-209 (2002)
9. S. Paczesny, O. I. Krijanovski, T. M. Braun, S. W. Choi, S. G. Clouthier, R. Kuick, D. E. Misek, K. R. Cooke, C. L. Kitko, A. Weyand, D. Bickley, D. Jones, J. Whitfield, P. Reddy, J. E. Levine, S. M. Hanash and J. L. Ferrara: A biomarker panel for acute graft-versus-host disease. *Blood*, 113, 273-278 (2009)
10. A. J. Rai, R. M. Kamath, W. Gerald and M. Fleisher: Analytical validation of the GeXP analyzer and design of a workflow for cancer-biomarker discovery using multiplexed gene-expression profiling. *Anal Bioanal Chem*, 393, 1505-1511 (2009)
11. F. Nigsch, A. Bender, B. van Buuren, J. Tissen, E. Nigsch and J. B. Mitchell: Melting point prediction employing k-nearest neighbor algorithms and genetic parameter optimization. *J Chem Inf Model*, 46, 2412-2422 (2006)
12. J. B. Marteau, S. Mohr, M. Pfister and S. Visvikis-Siest: Collection and storage of human blood cells for mRNA expression profiling: a 15-month stability study. *Clin Chem*, 51, 1250-1252 (2005)
13. L. Rainen, U. Oelmueller, S. Jurgensen, R. Wyrich, C. Ballas, J. Schram, C. Herdman, D. Bankaitis-Davis, N. Nicholls, D. Trollinger and V. Tryon: Stabilization of mRNA expression in whole blood samples. *Clin Chem*, 48, 1883-1890 (2002)
14. J. Aaroe, T. Lindahl, V. Dumeaux, S. Saebo, D. Tobin, N. Hagen, P. Skaane, A. Lonneborg, P. Sharma and A. L. Borresen-Dale: Gene expression profiling of peripheral blood cells for early detection of breast cancer. *Breast Cancer Res*, 12, R7 (2010)
15. M. R. Kuzman, V. Medved, J. Terzic and D. Krainc: Genome-wide expression analysis of peripheral blood identifies candidate biomarkers for schizophrenia. *J Psychiatr Res*, 43, 1073-1077 (2009)
16. M. Han, C. T. Liew, H. W. Zhang, S. Chao, R. Zheng, K. T. Yip, Z. Y. Song, H. M. Li, X. P. Geng, L. X. Zhu, J. J. Lin, K. W. Marshall and C. C. Liew: Novel blood-based, five-gene biomarker set for the detection of colorectal cancer. *Clin Cancer Res*, 14, 455-460 (2008)
17. M. Qin, D. Y. Wang, F. Huang, K. Nie, M. Qu, M. Wang, Y. L. Shu and X. J. Ma: Detection of pandemic influenza A H1N1 virus by multiplex reverse transcription-PCR with a GeXP analyzer. *J Virol Methods*, 168, 255-258 (2010)
18. M. A. Nagel, D. Gilden, T. Shade, B. Gao and R. J. Cohrs: Rapid and sensitive detection of 68 unique varicella zoster virus gene transcripts in five multiplex reverse transcription-polymerase chain reactions. *J Virol Methods*, 157, 62-68 (2009)
19. P. Muller, E. Warr, B. J. Stevenson, P. M. Pignatelli, J. C. Morgan, A. Steven, A. E. Yawson, S. N. Mitchell, H. Ranson, J. Hemingway, M. J. Paine and M. J. Donnelly: Field-caught permethrin-resistant *Anopheles gambiae* overexpress CYP6P3, a P450 that metabolises pyrethroids. *PLoS Genet*, 4, e1000286 (2008)
20. M. Ismail, G. Al-Naqeeb, W. A. Mamat and Z. Ahmad: Gamma-oryzanol rich fraction regulates the expression of antioxidant and oxidative stress related genes in stressed rat's liver. *Nutr Metab (Lond)*, 7, 23 (2010)
21. Q. R. Chen, G. Vansant, K. Oades, M. Pickering, J. S. Wei, Y. K. Song, J. Monforte and J. Khan: Diagnosis of the small round blue cell tumors using multiplex polymerase chain reaction. *J Mol Diagn*, 9, 80-88 (2007)
22. R. Li, H. Wang, B. N. Bekele, Z. Yin, N. P. Caraway, R. L. Katz, S. A. Stass and F. Jiang: Identification of putative oncogenes in lung adenocarcinoma by a comprehensive functional genomic approach. *Oncogene*, 25, 2628-2635 (2006)
23. J. P. Pinto, P. Ramos, S. F. de Almeida, S. Oliveira, L. Breda, M. Michalak, G. Porto, S. Rivella and M. de Sousa: Protective role of calreticulin in HFE hemochromatosis. *Free Radic Biol Med*, 44, 99-108 (2008)
24. L. Zou, M. Jaramillo, D. Whaley, A. Wells, V. Panchapakesa, T. Das and P. Roy: Profilin-1 is a negative regulator of mammary carcinoma aggressiveness. *Br J Cancer*, 97, 1361-1371 (2007)
25. Z. Ding, A. Lambrechts, M. Parepally and P. Roy: Silencing profilin-1 inhibits endothelial cell proliferation, migration and cord morphogenesis. *J Cell Sci*, 119, 4127-4137 (2006)
26. W. Engstrom, A. Ward and K. Moorwood: The role of scaffold proteins in JNK signalling. *Cell Prolif*, 43, 56-66 (2010)

27. D. Kanojia, M. Garg, S. Gupta, A. Gupta and A. Suri: Sperm-associated antigen 9, a novel biomarker for early detection of breast cancer. *Cancer Epidemiol Biomarkers Prev*, 18, 630-639 (2009)
28. M. Garg, D. Kanojia, S. Salhan, S. Suri, A. Gupta, N. K. Lohiya and A. Suri: Sperm-associated antigen 9 is a biomarker for early cervical carcinoma. *Cancer*, 115, 2671-2683 (2009)
29. V. Gerke, C. E. Creutz and S. E. Moss: Annexins: linking Ca<sup>2+</sup> signalling to membrane dynamics. *Nat Rev Mol Cell Biol*, 6, 449-461 (2005)
30. S. H. Ahn, H. Sawada, J. Y. Ro and G. L. Nicolson: Differential expression of annexin I in human mammary ductal epithelial cells in normal and benign and malignant breast tissues. *Clin Exp Metastasis*, 15, 151-156 (1997)
31. T. Masaki, M. Tokuda, M. Ohnishi, S. Watanabe, T. Fujimura, K. Miyamoto, T. Itano, H. Matsui, K. Arima, M. Shirai, T. Maeba, K. Sogawa, R. Konishi, K. Taniguchi, Y. Hatanaka, O. Hatase and M. Nishioka: Enhanced expression of the protein kinase substrate annexin in human hepatocellular carcinoma. *Hepatology*, 24, 72-81 (1996)
32. G. Shiota, D. B. Rhoads, T. C. Wang, T. Nakamura and E. V. Schmidt: Hepatocyte growth factor inhibits growth of hepatocellular carcinoma cells. *Proc Natl Acad Sci U S A*, 89, 373-377 (1992)
33. H. Tsubouchi, Y. Niitani, S. Hirono, H. Nakayama, E. Gohda, N. Arakaki, O. Sakiyama, K. Takahashi, M. Kimoto, S. Kawakami and et al.: Levels of the human hepatocyte growth factor in serum of patients with various liver diseases determined by an enzyme-linked immunosorbent assay. *Hepatology*, 13, 1-5 (1991)
34. J. Benn and R. J. Schneider: Hepatitis B virus HBx protein activates Ras-GTP complex formation and establishes a Ras, Raf, MAP kinase signaling cascade. *Proc Natl Acad Sci U S A*, 91, 10350-10354 (1994)
35. F. Anatelli, S. T. Chuang, X. J. Yang and H. L. Wang: Value of glypican 3 immunostaining in the diagnosis of hepatocellular carcinoma on needle biopsy. *Am J Clin Pathol*, 130, 219-223 (2008)
36. Y. Ito, A. Ando, H. Ando, J. Ando, Y. Saijoh, H. Inoko and H. Fujimoto: Genomic structure of the spermatid-specific hsp70 homolog gene located in the class III region of the major histocompatibility complex of mouse and man. *J Biochem*, 124, 347-353 (1998)
37. Y. Chen, G. Stamatoyannopoulos and C. Z. Song: Down-regulation of CXCR4 by inducible small interfering RNA inhibits breast cancer cell invasion in vitro. *Cancer Res*, 63, 4801-4804 (2003)
38. D. Uchida, N. M. Begum, A. Almofti, K. Nakashiro, H. Kawamata, Y. Tateishi, H. Hamakawa, H. Yoshida and M. Sato: Possible role of stromal-cell-derived factor-1/CXCR4 signaling on lymph node metastasis of oral squamous cell carcinoma. *Exp Cell Res*, 290, 289-302 (2003)
39. C. C. Schimanski, R. Bahre, I. Gockel, A. Muller, K. Frerichs, V. Horner, A. Teufel, N. Simiantonaki, S. Biesterfeld, T. Wehler, M. Schuler, T. Achenbach, T. Junginger, P. R. Galle and M. Moehler: Dissemination of hepatocellular carcinoma is mediated via chemokine receptor CXCR4. *Br J Cancer*, 95, 210-217 (2006)

**Abbreviations:** HCC: hepatocellular carcinoma; KNN: K nearest neighbors classifier; HBV: hepatitis B; LC: liver cirrhosis; RT: reverse transcription; A.U.: arbitrary units; AUC: area under the curve.

**Key Words:** Peripheral blood, GeXP, Hepatocellular carcinoma, Cross validation, Genes expression.

**Send correspondence to:** YP Tian, Department of Clinical Biochemistry, Chinese PLA General Hospital, 28 Fu-Xing Road, Beijing, China, Tel: 86-10-66939374, E-mail: tianyp61@gmail.com

<http://www.bioscience.org/current/vol4E.htm>