Received: 6 January 2018 | Accepted: 12 April 2018

DOI: 10.1111/liv.13864

CANCER

WILEY Liver INTERNATIONAL

A qualitative signature for early diagnosis of hepatocellular carcinoma based on relative expression orderings

Lu Ao[1] | Zimei Zhang[1] | Qingzhou Guan[1] | Yating Guo[1] | You Guo[1] | Jiahui Zhang[1] | Xingwei Lv[1] | Haiyan Huang[1] | Huarong Zhang[1] | Xianlong Wang[1,2] | Zheng Guo[1,2,3]

[1]Department of Bioinformatics, Key Laboratory of Ministry of Education for Gastrointestinal Cancer, School of Basic Medical Sciences, Fujian Medical University, Fuzhou, China

[2]Key Laboratory of Medical Bioinformatics, Fujian Province, Fuzhou, China

[3]Fujian Key Laboratory of Tumor Microbiology, Fujian Medical University, Fuzhou, China

Correspondence
Zheng Guo, Department of Bioinformatics, Key Laboratory of Ministry of Education for Gastrointestinal Cancer, School of Basic Medical Sciences, Fujian Medical University, Fuzhou, China.
Email: guoz@ems.hrbmu.edu.cn

Funding information
This work was supported by the National Natural Science Foundation of China [Grant Nos. 81602738, 81372213, 81572935 and 21534008], Young and Middle-Aged Backbone Training Project in the Health System of Fujian Province (Grant No. 2017-ZQN-56), Outstanding Youth Scientific Research Personnel Training Program in Fujian Province University (Grant No. 2017B018), the Joint Scientific and Technology Innovation Fund of Fujian Province (Grant No: 2016Y9044) and Chinese National Undergraduate Training Programs for Innovation and Entrepreneurship (Grant No: 201610392090).

Handling Editor: Chun-Jen Liu

Abstract

Background & Aims: Currently, using biopsy specimens to confirm suspicious liver lesions of early hepatocellular carcinoma are not entirely reliable because of insufficient sampling amount and inaccurate sampling location. It is necessary to develop a signature to aid early hepatocellular carcinoma diagnosis using biopsy specimens even when the sampling location is inaccurate.

Methods: Based on the within-sample relative expression orderings of gene pairs, we identified a simple qualitative signature to distinguish both hepatocellular carcinoma and adjacent non-tumour tissues from cirrhosis tissues of non-hepatocellular carcinoma patients.

Results: A signature consisting of 19 gene pairs was identified in the training data sets and validated in 2 large collections of samples from biopsy and surgical resection specimens. For biopsy specimens, 95.7% of 141 hepatocellular carcinoma tissues and all (100%) of 108 cirrhosis tissues of non-hepatocellular carcinoma patients were correctly classified. Especially, all (100%) of 60 hepatocellular carcinoma adjacent normal tissues and 77.5% of 80 hepatocellular carcinoma adjacent cirrhosis tissues were classified to hepatocellular carcinoma. For surgical resection specimens, 99.7% of 733 hepatocellular carcinoma specimens were correctly classified to hepatocellular carcinoma, while 96.1% of 254 hepatocellular carcinoma adjacent cirrhosis tissues and 95.9% of 538 hepatocellular carcinoma adjacent normal tissues were classified to hepatocellular carcinoma. In contrast, 17.0% of 47 cirrhosis from non-hepatocellular carcinoma patients waiting for liver transplantation were classified to hepatocellular carcinoma, indicating that some patients with long-lasting cirrhosis could have already gained hepatocellular carcinoma characteristics.

Conclusions: The signature can distinguish both hepatocellular carcinoma tissues and tumour-adjacent tissues from cirrhosis tissues of non-hepatocellular carcinoma

Abbreviations: HCC, hepatocellular carcinoma; REOs, relative expression orderings.

Lu Ao, Zimei Zhang, and Qingzhou Guan contributed equally to this work.

patients even using inaccurately sampled biopsy specimens, which can aid early diagnosis of hepatocellular carcinoma.

KEYWORDS
biopsy, cirrhosis, early diagnosis, hepatocellular carcinoma, signature

1 | INTRODUCTION

Liver cancer is the second most common cause of death from cancer and over 90% of cases are hepatocellular carcinoma (HCC). The diagnosis of HCC at an early stage is very important to improve disease prognosis.[1-3] Unfortunately, only about 30% of HCC in developed countries are diagnosed at an early stage and the figure is lower in developing countries.[4] At present, the diagnosis of HCC mainly depends on imaging techniques and serum biomarkers.[5] The sensitivities of imaging techniques, including ultrasonography, computed tomography and magnetic resonance imaging, vary greatly depending on the lesion size and operator experience. For histologically well-differentiated tumours with diameter smaller than 2 cm, usually referring to as early HCC, their sensitivities are all below 50% even for experienced pathologists.[5,6] The diagnostic sensitivity of serum markers, α-foetoprotein, is around 60% for early HCC patients.[7]

Thus, tissue biopsy becomes a necessary method for the early diagnosis of HCC. However, because the biopsy location may be inaccurate,[8] the false-negative rate of diagnosis based on biopsy specimens, with inaccurately sampled adjacent non-tumour tissues (cirrhosis or normal) of HCC, is about 30% and it increases to 50% in small biopsy specimens.[9,10] Therefore, it is necessary to develop signatures to distinguish HCC from tissues of cirrhosis patients based on a minimum biopsy specimen even when the biopsy location is inaccurate, which would be possible because the adjacent non-tumour liver tissues of HCC patients might have some molecular characteristics of HCC.[11-13] However, previously reported diagnostic signatures, such like the 2 transcriptional signatures reported by Chuma et al[14] and Jia et al,[15] all took tumour-adjacent non-tumour liver tissues as the normal samples to obtain the signature genes. Thus, these signatures cannot classify inaccurately sampled HCC adjacent non-tumour tissues (cirrhosis or normal) to HCC. Another major limitation of the previously reported diagnostic signatures is that their applications are based on risk scores summarized from signature genes' quantitative expression measurements,[16,17] which lack robustness for clinical applications because of large measurement batch effects[18] and quality uncertainties of clinical samples.[19-21]

Fortunately, the within-sample relative expression orderings (REOs) of genes, which are the qualitative transcriptional characteristics of samples, are robust against to experimental batch effects and disease signatures based on REOs can be directly applied to samples at the individualized level.[22-26] Besides, we have reported that the within-sample REOs of genes are highly robust against to varied proportions of the tumour epithelial cell in tumour tissues sampled from different tumour locations of the same patient,[20] partial RNA degradation during specimen storage and preparation[19] and amplification bias for minimum specimens,[21] which are common factors that can lead to failure of a quantitative transcriptional signature in clinical applications. Therefore, it is worthy to exploit the within-sample REOs to identify a robust qualitative signature for early diagnosis of HCC using minimum biopsy specimens.

Key points

- This study used the within-sample relative expression orderings (REOs) to develop a signature to aid early HCC diagnosis using biopsy specimens.
- The signature consisting of 19 gene pairs can distinguish both HCC tissues and tumour-adjacent tissues from cirrhosis tissues of non-HCC patients even using inaccurately sampled biopsy specimens.
- For biopsy specimens, 95.7% of 141 HCC tissues and all of 108 cirrhosis tissues of non-HCC patients were correctly classified.
- For surgical resection specimens, 99.7% of 733 HCC specimens were correctly classified, while 17.0% of 47 cirrhosis from non-HCC patients waiting for liver transplantation were classified to HCC.

In this study, we identified a qualitative signature based on the REOs of 19 gene pairs for early diagnosis of HCC. Then, we validated that the signature can accurately discriminate HCC tissues, including HCC adjacent non-tumour (cirrhosis or normal) liver tissues, from cirrhosis tissue of non-HCC patients in both surgical resection and biopsy samples. The results together suggested that the signature could aid early diagnosis of HCC even when the sampling location of biopsy specimen is inaccurate. Besides, through the analysis for patients with advanced cirrhosis of the liver terminal waiting for liver transplantation, we provided primary evidence that the signature might be able to identify cirrhosis patients at high risk of HCC.

2 | MATERIALS AND METHODS

2.1 | Data sources and data preprocessing

Multiple gene expression profiles were collected from Gene Expression Omnibus repository (GEO), as described in Table 1. Especially, 81 HCC

TABLE 1 Description of data sets used in this study

Data set	Platform	HCC	CoHCC	CHCC	Adjacent normal	Reference
Data sets used for identification of the qualitative signature						
GSE14323	GPL571	38	41	–	–	27
GSE15654-T[a]	GPL8432	–	108	–	–	28
GSE14520	GPL570	225	–	–	–	29
GSE63898	GPL13667	228	–	–	–	30
Total		491	149			
Data sets from biopsy used for evaluating the performance of the qualitative signature						
GSE15654-V[a]	GPL8432	–	108	–	–	28
GSE54236[a]	GPL6480	81	–	80	–	31
GSE64041[a]	GPL6244	60	–	–	60	32
Total		141	108	80	60	
Data sets from surgical resection used for evaluating the performance of the qualitative signature						
GSE17967	GPL571	–	47	16	–	17
GSE6764	GPL570	35	–	10	–	16
GSE17548	GPL570	17	–	20	–	33
GSE41804	GPL570	20	–	–	20	34
GSE62232	GPL570	81			10	35
GSE25097	GPL10687	268	–	40	243	36
GSE36376	GPL10558	240	–	–	193	37
GSE39791	GPL10558	72	–	–	72	38
GSE63898	GPL13667	–	–	168	–	30
Total		733	47	254	538	
RNA-Seq data for evaluating the performance of the qualitative signature						
TCGA	HTSeq- FPKM	355			42	39

CoHCC and CHCC denote cirrhosis tissues of non-HCC patients and adjacent cirrhosis tissues of HCC patients respectively.
[a]Samples collected by biopsy.

samples in the data set GSE54236 were collected from biopsy specimens with 500 ng total RNA for each sample measured by the Agilent platform, and 60 HCC samples in the data set GSE64041 were also collected from biopsy specimens with 250 ng total RNA for each sample measured by the Affymetrix platform. The 216 cirrhosis samples of non-HCC patients in the data set GSE15654 were obtained from small biopsies followed by formalin fixation (typically 10 × 1 mm pieces of tissue).[33] Notably, HCC samples denote cancerous tissue samples from HCC patients. Non-HCC cirrhosis samples denote the cirrhosis tissue samples from cirrhosis patients without HCC. HCC adjacent non-tumour samples denote the tumour-adjacent cirrhosis or normal tissue samples from HCC patients.

The Robust Multi-array Average algorithm[40] was used to process the data measured by the Affymetrix platform for background adjustment without inter-sample normalization. If several probesets were mapped to a gene, the expression value for the gene was defined as the arithmetic mean of the values of the multiple probesets (on the log2 scale). As to the data sets measured by the Illumina and Agilent platforms, we directly used the processed expression data. For the RNA-Seq data, the FPKM (Fragments Per Kilobase of transcript per Million fragments mapped)[41] values were directly download from TCGA. Then, the Ensembl gene IDs corresponding to the unique Entrez gene IDs of protein coding genes were used for further analysis.

2.2 | Identification of the qualitative REO-based diagnostic signature

For a gene pair, gene a and b with expression levels of G_a and G_b, respectively, if the REO pattern ($G_a > G_b$ or $G_a < G_b$) is kept in more than 85% of HCC samples in the training data, and reversed in more than 85% cirrhosis samples of non-HCC patients, then this gene pair is defined as reversal gene pair between the 2 types of samples. The rank difference for each reversal gene pair in each HCC or cirrhosis sample of non-HCC is calculated as follow:

$$R_{ij} = |R_i - R_j|,$$

R_i and R_j represent the ranks of gene i and j in a sample, respectively, and R_{ij} is the absolute rank difference between the 2 genes. Let mean [R_{ij}(cirr)] and mean [R_{ij}(hcc)] represent the means of absolute rank differences of the reversal gene pair (i, j) in all cirrhosis samples of non-HCC and all HCC samples respectively. Then, the geometric mean of the mean [R_{ij}(cirr)]and the mean [R_{ij}(hcc)] were calculated to evaluate

the reversal degree of the gene pair between HCC and cirrhosis of non-HCC:

$$\text{avg}R_{ij} = \sqrt{\text{mean}[R_{ij}(\text{cirr})] \times \text{mean}[R_{ij}(\text{hcc})]}.$$

The larger this geometric mean, the larger the reversal degree of the REO of the gene pair is between the 2 types of samples. All reversal gene pairs were sorted in a descending order according to their reversal degrees. The gene pair with the largest reversal degree was selected as the seed, and then a forward selection procedure was used to search an odd number of optimal subset of the reversal gene pairs to achieve the highest classification accuracy based on the majority voting rule. For a given sample, when more than a half gene pairs in the signature show the REOs for HCC, the sample is classified to HCC; otherwise, it is classified to cirrhosis of non-HCC.

2.3 | Performance evaluation

Hepatocellular carcinoma samples and cirrhosis samples of non-HCC from different data sets were directly pooled together. The sensitivity, specificity and accuracy were used to evaluate the performance of the signature. Here, the sensitivity was defined as the proportion of correctly identified HCC samples in all HCC samples and the specificity was defined as the proportion of correctly identified cirrhosis samples of non-HCC in all non-HCC samples. The accuracy was defined as the proportion of correctly identified samples of all HCC and non-HCC samples.

The receiver operating characteristic (ROC) curves were calculated as following[42]: (i) Among the 19 pairs of the signature, the number of gene pairs characterizing the REO patterns of the HCC was counted for each sample; (ii) Each sample was classified to be HCC if the above count was greater or equal to the voting threshold for HCC, ranging from 1 to 19; (iii) At each threshold, the true positive rates (sensitivity) and the false positive rates (1-specificity) were calculated. The sensitivity was the proportion of actual HCC which were correctly classified as HCC, and the specificity was the proportion of actual non-HCC which are correctly classified as non-HCC; (iv) The area under curve (AUC) was calculated with the nonparametric Hanley-McNeil algorithm[43] and 95% confidence intervals for AUC was determined using an approximate normal distribution.

2.4 | Statistical analysis

All statistical analyses were performed using the R 3.1.1 language (http://www.r-project.org/).

3 | RESULTS

3.1 | Identification of the qualitative diagnostic signature

The flow chart of this study is described in Figure 1. Firstly, we identified a total of 35,987,367 gene pairs with the same REOs in more than 85% of the 491 HCC samples collected from 3 data sets of GSE14323, GSE14520 and GSE63898. Similarly, we identified 24,741,470 gene pairs with the same REOs in more than 85% of the 149 cirrhosis samples of non-HCC collected from 2 data sets of GSE14323 and GSE15654. Here, because of the limited cirrhosis samples of non-HCC patients, 216 cirrhosis samples in GSE15654 were divided into 2 parts according to the GSM series numbers of samples: the first 108 cirrhosis samples, denoted as the GSE15654-T subset, were used to develop the qualitative diagnostic signature and the remained 108 cirrhosis samples, denoted as the GSE15654-V subset, were used to validate the signature.

A total 72 gene pairs showed reversal REOs between HCC tissues and cirrhosis tissues of non-HCC. Then, the 72 gene pairs were sorted in a descending order according to their reversal degrees (see Materials and Methods) between HCC and cirrhosis of non-HCC in the training data. According to the process described in the Materials and Methods, 19 gene pairs were selected from the 72 gene pairs as the diagnostic signature, denoted as the 19-gene-pair (shown in Figure 2 and Table 2). With this signature, based on the majority voting rule, 99.8% HCC samples and 99.3% cirrhosis samples of non-HCC patients in the training data sets were correctly classified. The detailed classification performance of the signature in each of the training data set was shown in Table S1.

3.2 | Validation of the diagnostic signature in independent data sets

Then the 19-gene-pair was validated in multiple data sets of biopsy and surgical resection samples respectively.

For biopsy samples in the data set GSE64041 with 250 ng RNA for each sample, 100.0% of the 60 HCC samples were correctly classified to HCC by the 19-gene-pair. In the data set GSE54236 with 500 ng RNA for each sample, 92.6% of the 81 HCC samples were classified to HCC. In other word, 95.7% of 141 HCC tissues were correctly classified. The accuracy was 97.6% and the AUC was 0.9999 (95% CI = 0.9705-1; shown in Figure 3A). Meanwhile, 100.0% of the 108 cirrhosis biopsy samples of non-HCC patients in the GSE15654-V subset were correctly classified. These results validated that the signature could discriminate HCC from cirrhosis samples of non-HCC. Moreover, 100.0% of the 60 HCC adjacent normal biopsy tissues in the data set GSE64041 and 77.5% of the 80 HCC adjacent cirrhosis biopsy tissues in the data set GSE54236 were classified to HCC. The results validated that, even using the inaccurately sampled biopsy specimens, the 19-gene-pair could classify most of tumour-adjacent tissues of HCC patients to HCC. For surgical resection samples, as shown in Table 3, 99.7% of the 733 HCC samples integrated from 7 data sets were correctly classified. Moreover, 96.1% of the 254 HCC adjacent cirrhosis samples and 95.9% of the 538 HCC adjacent normal samples were classified to HCC (Table S2). The accuracy was 98.8% and the AUC was 0.9452 (95% CI = 0.8803-1; shown in Figure 3B). The results suggested that the 19-gene-pair could identify most of adjacent non-tumour liver tissues from HCC patients to HCC. On the contrary, 83.0% of the 47 patients with advanced cirrhosis of the liver terminal waiting for liver transplantation[17] from the data set

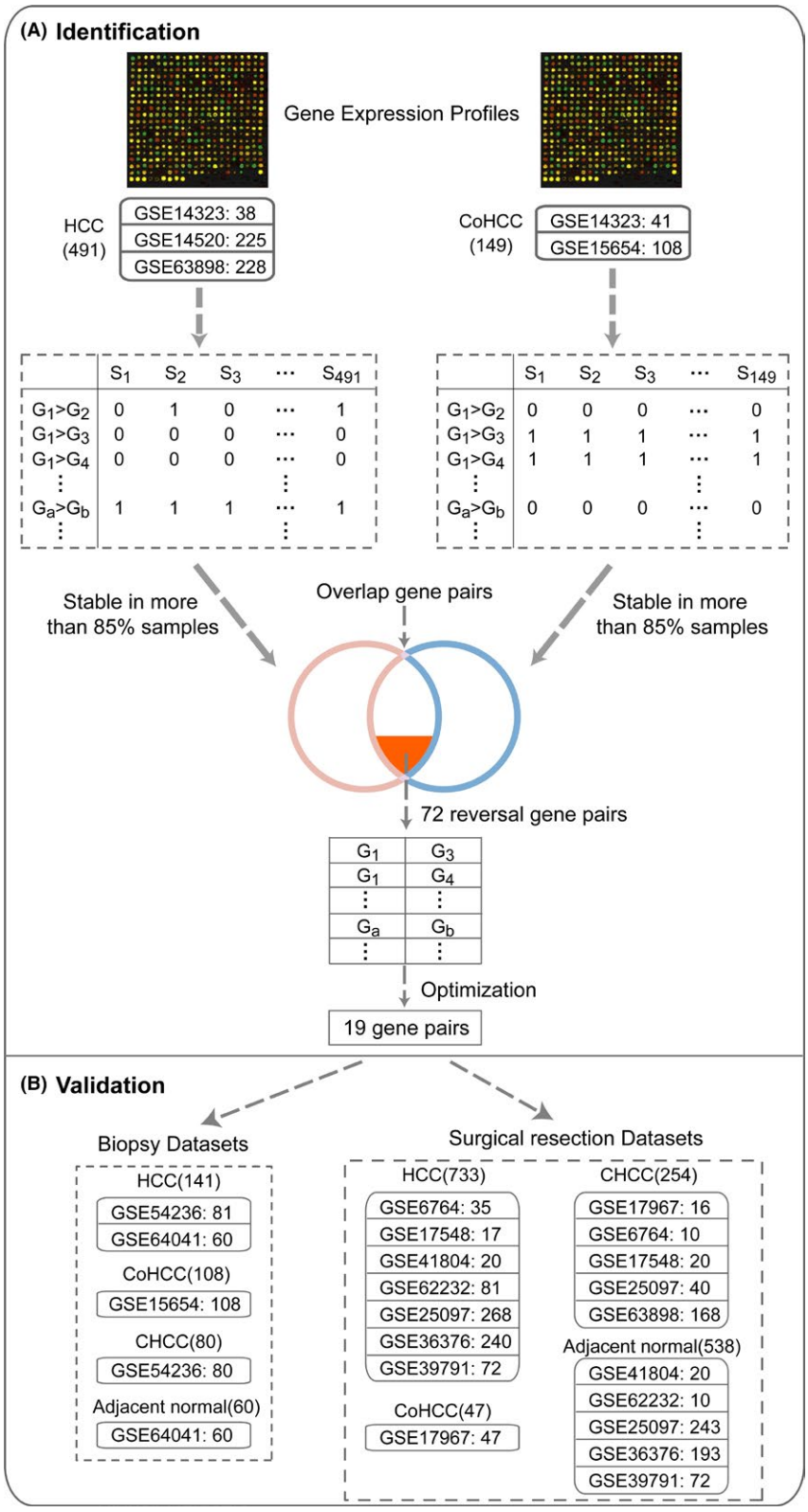

FIGURE 1 Workflow for identification and evaluation of the qualitative diagnostic signature. The workflow has 2 major analysis steps: A, identification and B, evaluation of the qualitative diagnostic signature in both biopsy and surgical resection samples. CoHCC and CHCC denote cirrhosis of non-HCC patients and adjacent cirrhosis of HCC patients respectively. Pink circle represents gene pairs with highly stable REOs in HCC samples and blue circle represents gene pairs with highly stable REOs in cirrhosis of non-HCC samples. Red denotes the 72 gene pairs with reversal REOs between these 2 types of samples and 19 gene pairs with the highest prediction accuracy in the training data were selected as the qualitative signature

GSE17967 were classified to non-HCC, whereas 17.0% of these samples were classified as HCC. The result indicated that the signature might be able to identify cirrhosis patients at high risk of HCC because it is very possible that a certain percentage of the patients with long-lasting cirrhosis could have already gained some characteristics of HCC.

Notably, the 19 gene-pair signature could also be validated using the RNA-Seq data set of HCC from TCGA: all the 355 HCC samples and 42 HCC adjacent non-tumour tissues were correctly classified to HCC. However, because we could find RNA-Seq data of cirrhosis tissue samples from non-HCC patients in neither TCGA nor GEO, we were unable to test the signature on cirrhosis samples. Notably,

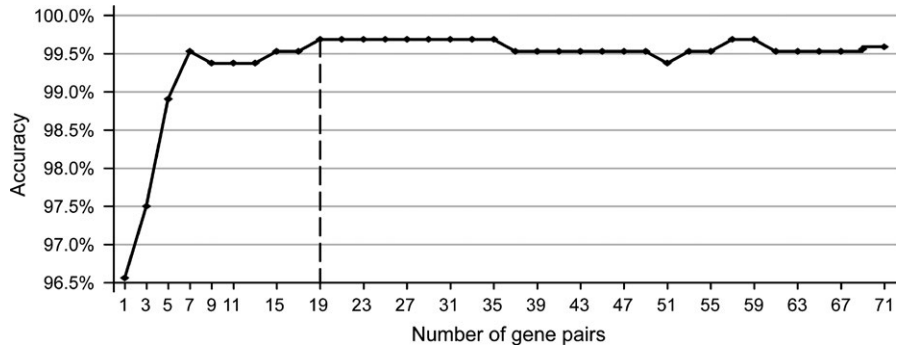

FIGURE 2 The accuracy of top-ranked gene pairs from the 72 gene pairs in the training data. The 72 reversal gene pairs were sorted in a descending order according to their reversal degrees between HCC tissues and cirrhosis tissues of non-HCC in the training data. The 19 gene pairs from 72 reversal gene pairs achieved the highest classification accuracy based on the majority voting rule and were selected as the qualitative signature

TABLE 2 The 19-gene-pair signature for early diagnosis of HCC

Signature	Gene A	Gene B
pair1	VAT1	CHST4
pair2	HMGN1	PHF11
pair3	GLUD2	PROM1
pair4	TMEM38B	AGO3
pair5	RRAGD	AGO3
pair6	KHDRBS3	AGO3
pair7	PCOLCE2	PTBP3
pair8	HNF1A	MAPRE3
pair9	NKRF	RHBDF1
pair10	LSM5	AGO3
pair11	ACTR5	CTF1
pair12	CHAF1A	CTF1
pair13	CDCA4	PROM1
pair14	MOSPD2	AGO3
pair15	SMC4	AGO3
pair16	LIN7C	PRF1
pair17	TSNAX	MCL1
pair18	TBCE	AGO3
pair19	GADD45GIP1	AGO3

Gene A has a higher expression level than Gene B in HCC patients compared with cirrhosis tissues of non-HCC patients.

among the 355 HCC samples from TCGA, 104 samples had the history of alcohol consumption, 101 samples had the history of hepatitis B infection, 49 samples had the history of Hepatitis C infection; 335 samples have the stage information, 165 patients with stage I, 80 patients with stage II, 85 patients with stage III and 5 patients with stage IV. Regardless of the HCC aetiology and clinic stage, all of 355 HCC samples were correctly classified to HCC. The HCC aetiology or clinic stage status does not affect the validation results using the GEO data set either, as shown in the Data S1. The results demonstrated that the signature was robust to clinicopathological variations.

4 | DISCUSSION

Liver biopsy plays an essential role in confirming a suspected liver lesion which do not show typical features of HCC by imaging or serum examination. In this study, we identified a robust qualitative signature, 19-gene-pair consisted of 29 genes, for early diagnosis of HCC, which can distinguish HCC and most of tumour-adjacent tissues from cirrhosis tissues of non-HCC patients. It means that, even using the inaccurately sampled biopsy specimens, this signature can still aid early diagnosis of HCC. A few genes in this signature, including HNF1A, SMC4, PROM1, HMGN1, CHST4, PHF11, AGO3 and MCL1, are well known HCC-related genes which might play a key role in the development from cirrhosis to HCC. For example, HNF1A is a liver-enriched transcription factor that is essential for maintaining liver function, which might play a suppressor's role during hepatocarcinogenesis.[44] Another gene, SMC4, associated with vascular invasion,[45] was previously suggested to be useful for the early detection of HCC,[46] Additionally, PROM1, was a marker closely correlated with hepatocarcinogenesis.[47] In addition, HMGN1,[44] CHAF1A,[48] CHST4,[49] AGO3[50] and MCL1[51] have also been reported to be closely correlated with HCC. The above results indicated that the genes of the signature might play important roles in the hepatocarcinogenesis and these functions need to be further studied in the further work.

Notably, the biopsy samples analysed in this study had relatively large amount of tissues to yield about 250-500 ng total RNA.[31,32] However, under many practical situations with a needle for biopsy, it is difficult to obtain biopsy specimens to yield sufficient quantity of RNA molecules for gene expression profiling or other molecular measurements.[9,52] Fortunately, as demonstrated in our recent study,[21] the REO-based signatures obtained from samples with sufficient total RNA can be robustly applied to samples with minimum specimens with as low as 150-250 pg total RNA for about 15-25 cancer cells. Therefore, it is highly possible that the 19-gene-pair could be applicable to biopsy samples with minimum sampling amounts.

In summary, with large collections of both biopsy and surgical resection samples, we identified and validated a robust qualitative signature consisting of 19 gene pairs for aiding early diagnosis of HCC.

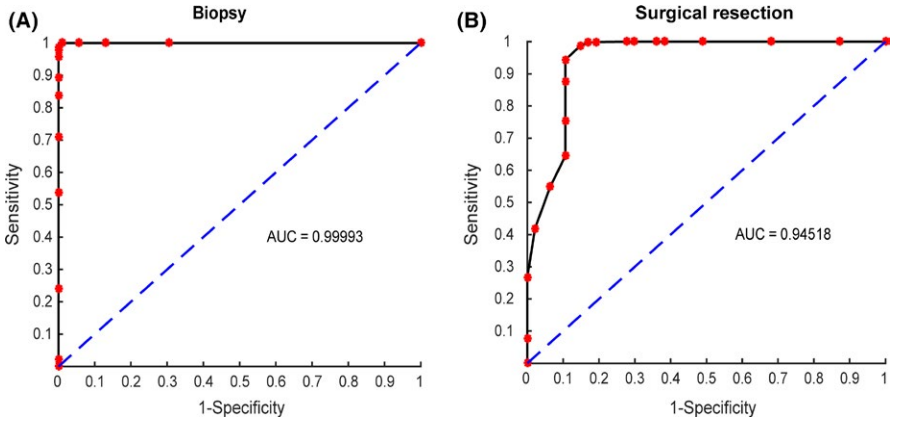

FIGURE 3 The receiver characteristic operating curves. Biopsy specimens with 141 HCC tissues and 108 cirrhosis tissues of non-HCC patients (A) and surgical resection specimens with 733 HCC tissues and 47 cirrhosis tissues of non-HCC patients (B)

TABLE 3 The performance of the signature in the validation data sets from surgical resection

Data set	Number (Sensitivity) of HCC samples	Number (Specificity) of cirrhosis samples of non-HCC
GSE6764	35 (100.0%)	—
GSE17548	17 (100.0%)	—
GSE41804	20 (100.0%)	—
GSE62232	81 (100.0%)	—
GSE25097	268 (100.0%)	—
GSE36376	240 (99.2%)	—
GSE39791	72 (100.0%)	—
GSE17967	—	47 (83.0%)
Total	733 (99.7%)	47 (83.0%)

The clinical value of the 19-gene-pair for early diagnosis of HCC is worthy to be further verified.

CONFLICT OF INTEREST

The authors do not have any disclosures to report.

ORCID

Lu Ao http://orcid.org/0000-0001-7378-4967

Zheng Guo http://orcid.org/0000-0003-4466-6026

REFERENCES

1. Nguyen MH, Keeffe EB. Screening for hepatocellular carcinoma. *J Clin Gastroenterol*. 2002;35:S86-S91.
2. Bertino G, Neri S, Bruno CM, et al. Diagnostic and prognostic value of alpha-fetoprotein, des-gamma-carboxy prothrombin and squamous cell carcinoma antigen immunoglobulin M complexes in hepatocellular carcinoma. *Minerva Med*. 2011;102:363-371.
3. Parikh A, Taouli B. Imaging of hepatocellular carcinoma: current concepts. *Recent Results Cancer Res*. 2013;190:33-55.
4. Kudo M. Japan's successful model of nationwide hepatocellular carcinoma surveillance highlighting the urgent need for global surveillance. *Liver Cancer*. 2012;1:141-143.
5. Fitzmorris P, Singal AK. Surveillance and diagnosis of hepatocellular carcinoma. *Gastroenterol Hepatol (N Y)*. 2015;11:38-46.
6. Yu NC, Chaudhari V, Raman SS, et al. CT and MRI improve detection of hepatocellular carcinoma, compared with ultrasound alone, in patients with cirrhosis. *Clin Gastroenterol Hepatol*. 2011;9:161-167.
7. Lok AS, Sterling RK, Everhart JE, et al. Des-gamma-carboxy prothrombin and alpha-fetoprotein as biomarkers for the early detection of hepatocellular carcinoma. *Gastroenterology*. 2010;138:493-502.
8. Forner A, Vilana R, Ayuso C, et al. Diagnosis of hepatic nodules 20 mm or smaller in cirrhosis: prospective validation of the noninvasive diagnostic criteria for hepatocellular carcinoma. *Hepatology*. 2008;47:97-104.
9. Forner A, Llovet JM, Bruix J. Hepatocellular carcinoma. *Lancet*. 2012;379:1245-1255.
10. Villanueva A, Minguez B, Forner A, Reig M, Llovet JM. Hepatocellular carcinoma: novel molecular approaches for diagnosis, prognosis, and therapy. *Annu Rev Med*. 2010;61:317-328.
11. Budhu A, Forgues M, Ye QH, et al. 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.
12. Hoshida Y, Villanueva A, Kobayashi M, et al. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N Engl J Med*. 2008;359:1995-2004.
13. Wei L, Lian B, Zhang Y, et al. Application of microRNA and mRNA expression profiling on prognostic biomarker discovery for hepatocellular carcinoma. *BMC Genom*. 2014;15(Suppl 1):S13.
14. Chuma M, Sakamoto M, Yamazaki K, et al. Expression profiling in multistage hepatocarcinogenesis: identification of HSP70 as a molecular marker of early hepatocellular carcinoma. *Hepatology*. 2003;37:198-207.
15. Jia HL, Ye QH, Qin LX, et al. Gene expression profiling reveals potential biomarkers of human hepatocellular carcinoma. *Clin Cancer Res*. 2007;13:1133-1139.
16. Wurmbach E, Chen YB, Khitrov G, et al. Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. *Hepatology*. 2007;45:938-947.

17. Archer KJ, Mas VR, David K, Maluf DG, Bornstein K, Fisher RA. Identifying genes for establishing a multigenic test for hepatocellular carcinoma surveillance in hepatitis C virus-positive cirrhotic patients. *Cancer Epidemiol Biomarkers Prev*. 2009;18:2929-2932.
18. Guan Q, Yan H, Chen Y, et al. Quantitative or qualitative transcriptional diagnostic signatures? A case study for colorectal cancer. *BMC Genom*. 2018;19:99.
19. Chen R, Guan Q, Cheng J, et al. Robust transcriptional tumor signatures applicable to both formalin-fixed paraffin-embedded and fresh-frozen samples. *Oncotarget*. 2017;8:6652-6662.
20. Cheng J, Guo Y, Gao Q, et al. Circumvent the uncertainty in the applications of transcriptional signatures to tumor tissues sampled from different tumor sites. *Oncotarget*. 2017;8:30265-30275.
21. Liu H, Li Y, He J, et al. Robust transcriptional signatures for low-input RNA samples based on relative expression orderings. *BMC Genom*. 2017;18:913.
22. Eddy JA, Sung J, Geman D, Price ND. Relative expression analysis for molecular cancer diagnosis and prognosis. *Technol Cancer Res Treat*. 2010;9:149-159.
23. Patil P, Bachant-Winner PO, Haibe-Kains B, Leek JT. Test set bias affects reproducibility of gene signatures. *Bioinformatics*. 2015;31:2318-2323.
24. Wang H, Sun Q, Zhao W, et al. Individual-level analysis of differential expression of genes and pathways for personalized medicine. *Bioinformatics*. 2015;31:62-68.
25. Guan Q, Chen R, Yan H, et al. Differential expression analysis for individual cancer samples based on robust within-sample relative gene expression orderings across multiple profiling platforms. *Oncotarget*. 2016;7:68909-68920.
26. Ao L, Song X, Li X, et al. An individualized prognostic signature and multiomics distinction for early stage hepatocellular carcinoma patients with surgical resection. *Oncotarget*. 2016;7:24097-24110.
27. Mas VR, Maluf DG, Archer KJ, et al. Genes involved in viral carcinogenesis and tumor initiation in hepatitis C virus-induced hepatocellular carcinoma. *Mol Med*. 2009;15:85-94.
28. Hoshida Y, Villanueva A, Sangiovanni A, et al. Prognostic gene expression signature for patients with hepatitis C-related early-stage cirrhosis. *Gastroenterology*. 2013;144:1024-1030.
29. Roessler S, Jia HL, Budhu A, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Can Res*. 2010;70:10202-10212.
30. Villanueva A, Portela A, Sayols S, et al. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. *Hepatology*. 2015;61:1945-1956.
31. Villa E, Critelli R, Lei B, et al. Neoangiogenesis-related genes are hallmarks of fast-growing hepatocellular carcinomas and worst survival. Results from a prospective study. *Gut*. 2016;65:861-869.
32. Makowska Z, Boldanova T, Adametz D, et al. Gene expression analysis of biopsy samples reveals critical limitations of transcriptome-based molecular classifications of hepatocellular carcinoma. *J Pathol Clin Res*. 2016;2:80-92.
33. Yildiz G, Arslan-Ergul A, Bagislar S, et al. Genome-wide transcriptional reorganization associated with senescence-to-immortality switch during human hepatocellular carcinogenesis. *PLoS One*. 2013;8:e64016.
34. Hodo Y, Honda M, Tanaka A, et al. Association of interleukin-28B genotype and hepatocellular carcinoma recurrence in patients with chronic hepatitis C. *Clin Cancer Res*. 2013;19:1827-1837.
35. Schulze K, Imbeaud S, Letouze E, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet*. 2015;47:505-511.
36. Tung EK, Mak CK, Fatima S, et al. Clinicopathological and prognostic significance of serum and tissue Dickkopf-1 levels in human hepatocellular carcinoma. *Liver Int*. 2011;31:1494-1504.
37. Lim HY, Sohn I, Deng S, et al. Prediction of disease-free survival in hepatocellular carcinoma by gene expression profiling. *Ann Surg Oncol*. 2013;20:3747-3753.
38. Kim JH, Sohn BH, Lee HS, et al. Genomic predictors for recurrence patterns of hepatocellular carcinoma: model derivation and validation. *PLoS Med*. 2014;11:e1001770.
39. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31:166-169.
40. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4:249-264.
41. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28:511-515.
42. Guo Y, Jiang W, Ao L, et al. A qualitative signature for predicting pathological response to neoadjuvant chemoradiation in locally advanced rectal cancers. *Radiother Oncol*. 2018; https://doi.org/10.1016/j.radonc.2018.01.010
43. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29-36.
44. Wang W, Hayashi Y, Ninomiya T, et al. Expression of HNF-1 alpha and HNF-1 beta in various histological differentiations of hepatocellular carcinoma. *J Pathol*. 1998;184:272-278.
45. Zhou B, Chen H, Wei D, et al. A novel miR-219-SMC4-JAK2/Stat3 regulatory pathway in human hepatocellular carcinoma. *J Exp Clin Cancer Res*. 2014;33:55.
46. Zhou B, Yuan T, Liu M, et al. Overexpression of the structural maintenance of chromosome 4 protein is associated with tumor dedifferentiation, advanced stage and vascular invasion of primary liver cancer. *Oncol Rep*. 2012;28:1263-1268.
47. Yin S, Li J, Hu C, et al. CD133 positive hepatocellular carcinoma cells possess high capacity for tumorigenicity. *Int J Cancer*. 2007;120:1444-1450.
48. Xu M, Jia Y, Liu Z, et al. Chromatin assembly factor 1, subunit A (P150) facilitates cell proliferation in human hepatocellular carcinoma. *Onco Targets Ther*. 2016;9:4023-4035.
49. Gao F, Liang H, Lu H, et al. Global analysis of DNA methylation in hepatocellular carcinoma by a liquid hybridization capture-based bisulfite sequencing approach. *Clin Epigenetics*. 2015;7:86.
50. Kitagawa N, Ojima H, Shirakihara T, et al. Downregulation of the microRNA biogenesis components and its association with poor prognosis in hepatocellular carcinoma. *Cancer Sci*. 2013;104:543-551.
51. Jiang C, Long J, Liu B, Xie X, Kuang M. Mcl-1 is a novel target of miR-26b that is associated with the apoptosis induced by TRAIL in HCC cells. *Biomed Res Int*. 2015;2015:572738.
52. De Rienzo A, Yeap BY, Cibas ES, et al. Gene expression ratio test distinguishes normal lung from lung tumors in solid tissue and FNA biopsies. *J Mol Diagn*. 2014;16:267-272.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Ao L, Zhang Z, Guan Q, et al. A qualitative signature for early diagnosis of hepatocellular carcinoma based on relative expression orderings. *Liver Int*. 2018;38:1812–1819. https://doi.org/10.1111/liv.13864