Journal of
Translational Medicine

RESEARCH

Open Access

A robust qualitative transcriptional signature for the correct pathological diagnosis of gastric cancer

Haidan Yan[1†], Meifeng Li[1†], Longlong Cao[2], Haifeng Chen[3], Hungming Lai[1], Qingzhou Guan[1], Huxing Chen[1], Wenbin Zhou[4], Baotong Zheng[1], Zheng Guo[1,4*] and Chaohui Zheng[2*]

Abstract

Background: Currently, pathological examination of gastroscopy biopsy specimens is the gold standard for gastric cancer (GC) diagnosis. However, it has a false-negative rate of 10–20% due to inaccurate sampling locations and/or insufficient sampling amount. A signature should be developed to aid the early diagnosis of GC using biopsy specimens even when they are sampled from inaccurate locations.

Methods: We extracted a robust qualitative transcriptional signature, based on the within-sample relative expression orderings (REOs) of gene pairs, to discriminate both GC tissues and adjacent-normal tissues from non-GC gastritis, intestinal metaplasia and normal gastric tissues.

Results: A signature consisting of two gene pairs for GC diagnosis was identified and validated in data of both biopsy specimens and surgical resection specimens pooled from publicly available datasets measured by different laboratories with different platforms. For gastroscopy biopsy specimens, 96.20% of 79 non-GC tissues were correctly identified as non-GC, and 96.84% of 158 GC tissues and six of seven adjacent-normal tissues were correctly identified as GC. For surgical resection specimens, 98.37% of 2560 GC tissues and 97.28% of 221 adjacent-normal tissues were correctly identified as GC. Especially, 97.67% of the 257 GC patients at stage I were exactly diagnosed as GC. We additionally measured 21 GC tissues from seven different GC patients, each with three specimens sampled from three tumor locations with different proportions of the tumor epithelial cell. All these GC tissues were correctly identified as GC, even when the proportion of the tumor epithelial cell was as low as 14%.

Conclusions: The qualitative transcriptional signature can distinguish both GC and adjacent-normal tissues from normal, gastritis and intestinal metaplasia tissues of non-GC patients even using inaccurately sampled biopsy specimens, which can be applied robustly at the individual level to aid the early GC diagnosis.

Keywords: Gastric cancer, Gastritis, Gastroscopy biopsy, Diagnosis, Signature

*Correspondence: guoz@ems.hrbmu.edu.cn; wwkzch@163.com
†Haidan Yan and Meifeng Li have contributed equally to this work
[2] Department of Gastric Surgery, Fujian Medical University Union Hospital, No. 29 Xinquan Road, Fuzhou 350001, China
[4] Department of Systems Biology, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, China
Full list of author information is available at the end of the article

Background

Gastric cancer (GC) is one of the most frequent malignant tumors with a high mortality rate [1–3]. GC patients at early stage could benefit from surgical resection [4, 5]. However, only about 10–20% of GC patients are diagnosed at early stage [6, 7]. Currently, pathological examination based on gastroscopic biopsy tissue is still the most effective approach for confirming GC [8, 9]. However, the result of pathological examination for gastroscopic biopsy tissue depends on the skills and experiences of the endoscopists and pathologists [10–12]. The false-negative rate of GC diagnosis has been reported to be 10–20% [13–18]. Among the false-negative samples, 85.2% are at the early stage [19], and 71.4% are wrongly diagnosed as gastritis, ulcer or "suspicious lesion" [16]. Most of the false-negative samples (73%) are caused by inaccurate sampling locations and the remainder (27%) could be attributed to pathologist errors [16].

Therefore, it is vitally important to develop an objective molecular signature to complement the existing subjective diagnostic technique of histology, which could aid the pathologists to identify early GC even when the sampling location of gastroscopic biopsy tissue is inaccurate. It's possible because the GC adjacent-normal tissues might also gain some similar molecular characteristics of GC [20, 21]. However, most of the reported diagnostic signatures are identified using GC adjacent-normal tissues as the normal samples [22–24], which will make false-negative diagnosis when the location of gastroscopic biopsy tissue is inaccurate [13]. Another critical limitation of previously reported diagnostic signatures is that they are based on risk scores summarized from quantitative gene expression measurements of the signature genes [22, 23, 25], which are highly sensitive to measurement batch effects and lab differences and thus cannot be robustly applied to independent samples [26–28] even with data normalization [29]. Fortunately, it has been reported that the within-sample relative expression orderings (REOs) of genes are robust against experimental batch effects [30, 31]. Besides, we have shown that the within-sample REOs are robust even when the tumor tissues sampled from different tumor locations contain different proportions of the tumor epithelial cell [32] and partial RNA degradation during specimen preparation and storage [33], and the RNA amplification bias exists for minimum specimens. Notably, Zheng et al. have identified the within-sample REO of one pair of microRNA (hsa-miR-196a and hsa-miR-148a) as a qualitative GC diagnosis signature using GC and normal gastric mucosa samples [34]. However, the performance of this signature to identify gastritis, intestinal metaplasia and cancer adjacent-normal samples was not evaluated [34].

In this study, we aim at identifying a signature that can discriminate GC tissues, including the inaccurately sampled GC adjacent-normal tissues, from non-GC tissues including gastritis, intestinal metaplasia and normal gastric tissues. A signature consisting of two gene pairs was identified in the training data and validated in multiple datasets measured by different laboratories with different platforms, even when the proportion of the tumor epithelial cell was as low as 14%.

Materials and methods

Samples and data measurement

We measured 21 GC specimens from seven GC patients. For each patient, three specimens were sampled from three different tumor locations. The proportion (about 14%–93%) of the tumor epithelial cell was measured by pathological section analysis (see Table 1). The baseline characteristics of the seven GC patients were shown in Additional file 1: Table S1. All cancer specimens were collected from the operating room immediately after surgical resection and were fresh frozen for subsequent RNA extraction. This study was approved by the institutional review boards of all participating institutions, and written consent forms were obtained from all participants.

Total RNA was isolated from fresh frozen GC tissues using Trizol reagent (Invitrogen) according to the manufacture's protocol. The quality of RNA was assessed using Agilent 2200 TapeStation (Agilent technologies, US) to ensure high quality (RNA integrity number > 6). Then, 1–2 μg of total RNA was used for mRNA capture using NEBNextPolyA mRNA Magnetic Isolation Module and stranded RNA-seq libraries were constructed using a NEBNext Ultra Directional RNA Library Prep Kit. The 2 × 150 paired-end sequencing was performed on an Illumina HiSeqXten (Illumina, US). The resulting raw RNA-seq files (.fastq) were preprocessed using Trimmomatic [35], and reads were aligned to the reference genome (GRCh37) using hisat2 [36]. Finally, the reads

Table 1 The proportions of the tumor epithelial cell for GC tissues of each patient sampled from three different locations

Patient	Proportion 1 (%)	Proportion 2 (%)	Proportion 3 (%)
GC 1	23	79	53
GC 2	53	28	89
GC 3	27	73	93
GC 4	35	67	89
GC 5	88	37	14
GC 6	88	33	57
GC 7	15	74	47

per kilobase per million mapped reads (RPKM) values of genes were computed to represent the expression levels of genes using StringTie [37]. The data has been submitted to Gene Expression Omnibus (GEO, GSE116782).

Public data and preprocessing

Gene expression profiles of gastric tissues measured by the Affymetrix, Illumina or RNA-seq platform were collected from the GEO and The Cancer Genome Atlas (TCGA) data portal (http://tcga-data.nci.nih.gov/tcga/), as described in Table 2.

For the gene expression profiles measured by the Affymetrix platform, the raw data (.CEL files) was downloaded and preprocessed using the Robust Multi-array Average algorithm for background adjustment without quantile normalization [38]. For the gene expression profiles measured by the Illumina platform, the processed data was directly downloaded and used for the following analysis. For the gene expression profiles from TCGA detected by RNA-seq, the level 3 data was directly downloaded for our analysis.

For the array-based data, every probe ID was mapped to Entrez gene ID using the corresponding platform file. If multiple probes were mapped to a gene, the expression level of this gene was summarized as the arithmetic mean of the values of these probes.

Developing the diagnostic signature

The gene expression profiles of GC, normal and gastritis tissues in the training data were used to identify REO-based diagnostic signature (Table 2). First, we defined the stable REOs of gene pairs in a type of gastric tissues. The REO of a gene pair (i, j) is denoted as $Gi > Gj$ or $\mathrm{Gi} < Gj$ if the gene i has a higher or lower expression level than the gene j within a sample. The REO of a gene pair is defined as stable if the same REO kept in at least 99% of the samples. Furthermore, a gene pair (i, j) is defined as reversal if the REO of the gene pair is stable in both of two types of gastric tissues, but with different REO patterns ($Gi < Gj$ or $Gi > Gj$ in one type of tissues but $Gi > Gj$ or $Gi < Gj$ in the other type of tissues). Here, the stable gene pairs with the same REO pattern between normal samples and gastritis samples were defined as stable gene pairs of non-GC tissues. We then selected the reversal gene pairs between GC and non-GC tissue samples. These reversal gene pairs were the candidate qualitative REO-based diagnostic signatures. The absolute rank difference for every reversal gene pair in each of the GC or non-GC samples is calculated as follow:

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

where R_i and R_j represent the ranks of gene i and j in a sample, respectively.

Table 2 The publicly available datasets used in the study

Dataset	Platform	Normal	GI	GC	GC_adjacent
Training					
GSE54129	Afftmetrix GPL570	21[a]	–	111	–
GSE54043	Afftmetrix GPL570	5[a]	5[a]	–	–
GSE42252	Afftmetrix GPL570	–	–	5	–
GSE38749	Afftmetrix GPL570	–	–	15	–
GSE51725	Afftmetrix GPL570	–	–	8	–
GSE79973	Afftmetrix GPL570	–	–	10	–
GSE57303	Afftmetrix GPL570	–	–	70	–
GSE13911	Afftmetrix GPL570	–	–	38	–
GSE27411	Illumina GPL6255	–	18[a]	–	–
GSE28541	Illumina GPL13376	–	–	40	–
GSE29998	Illumina GPL6947	–	–	50	–
Total		26	23	347	–
Validation					
GSE5081	Afftmetrix GPL570	–	32[a]	–	–
GSE52138	Afftmetrix GPL96	–	–	13[a]	7[a]
GSE14210	Afftmetrix GPL571	–	–	145[a]	–
GSE106656	Afftmetrix GPL6244	–	21[a]	–	–
GSE34619	Afftmetrix GPL6244	10[a]	–	–	–
GSE29272	Afftmetrix GPL96	–	–	134	134
GSE34942	Afftmetrix GPL570	–	–	56	–
GSE22377	Afftmetrix GPL570	–	–	43	–
GSE19826	Afftmetrix GPL570	–	–	12	12
GSE35809	Afftmetrix GPL570	–	–	70	–
GSE51105	Afftmetrix GPL570	–	–	94	–
GSE15459	Afftmetrix GPL570	–	–	200	–
GSE62254	Afftmetrix GPL570	–	–	300	–
GSE13861	Illumina GPL6884	–	–	65	19
GSE38024	Illumina GPL10558	–	–	48	–
GSE26899	Illumina GPL6947	–	–	96	12
GSE26253	Illumina GPL8432	–	–	432	–
GSE84437	Illumina GPL6947	–	–	433	–
GSE26942	Illumina GPL6947	–	–	202	12
GSE60662	Agilent GPL13497	4[a]	12[a]	–	–
TCGA	RNA-seq	–	–	375	32
Total		14	65	2718	228

GI represent gastritis, gastritis adjacent-normal or intestinal metaplasia tissues. GC_adjacent represent the GC adjacent-normal tissues

[a] Denotes the samples were collected by gastroscopic biopsy

For a reversal gene pair (i, j), let *mean* [$R_{ij}(non)$] and *mean* [$R_{ij}(gc)$] denote the means of the absolute rank differences between gene i and gene j in non-GC tissue samples and GC tissue samples, respectively. Then, their geometric mean ($avgR_{ij}$) is calculated to evaluate the reversal degree of the gene pair between GC and non-GC tissue samples.

$$avgR_{ij} = \sqrt{mean[R_{ij}(non)] \times mean[R_{ij}(gc)]}$$

The larger the geometric mean for a reversal gene pair, the larger the reversal degree of the REO of the gene pair between GC and non-GC tissue samples. All reversal gene pairs were sorted in a descending order according to the geometric means.

Finally, we took the top *k* reversal gene pairs as a signature according to the reversal degrees of the identified reversal gene pairs, and a given sample was identified as GC tissue when at least a half of gene pairs in the signature exhibit the same REOs for GC; otherwise, it was identified as non-GC tissue. The signature achieved the highest classification accuracy in the training data was defined as GC diagnosis signature. All the analysis programs to develop the diagnostic signature were written using the R language (R 3.1.3). The program codes were shown in Additional file 2.

Performance evaluation

The sensitivity, specificity, accuracy and the area under curve (AUC) of the receiver operating characteristic (ROC) curves were used to evaluate the performance of the signature. The sensitivity was defined as the proportion of correctly identified GC samples in all GC samples. The specificity was defined as the proportion of correctly identified non-GC samples in all non-GC samples including normal tissues, gastritis adjacent-normal tissues and gastritis tissues. The accuracy was defined as the proportion of correctly identified samples of all GC and non-GC samples. Here, the nonparametric Hanley-McNeil algorithm was used to calculate the AUC value [39, 40] and 95% confidence intervals (CI) for AUC was computed using an approximate normal distribution.

Results

Identifying the diagnostic gene pair signature

The flowchart for the identification and validation of the qualitative diagnostic signature is described in Fig. 1.

Firstly, we identified gene pairs with an identical REO in at least 99% of 26 gastric normal samples, 23 gastritis samples and 347 GC samples, respectively, using the training data integrated from 11 datasets measured by the Affymetrix or Illumina platform (see Table 1). We found 32,483,417 overlapped gene pairs with the same stable REOs between the gastric normal and gastritis samples, among which six gene pairs had stable but reversal REOs in the GC tissues (Additional files 3 and 4), which were potential GC diagnostic signatures.

We then evaluated the reversal degrees of the six gene pairs with reversal REOs between the GC and non-GC samples including normal and gastritis samples in the

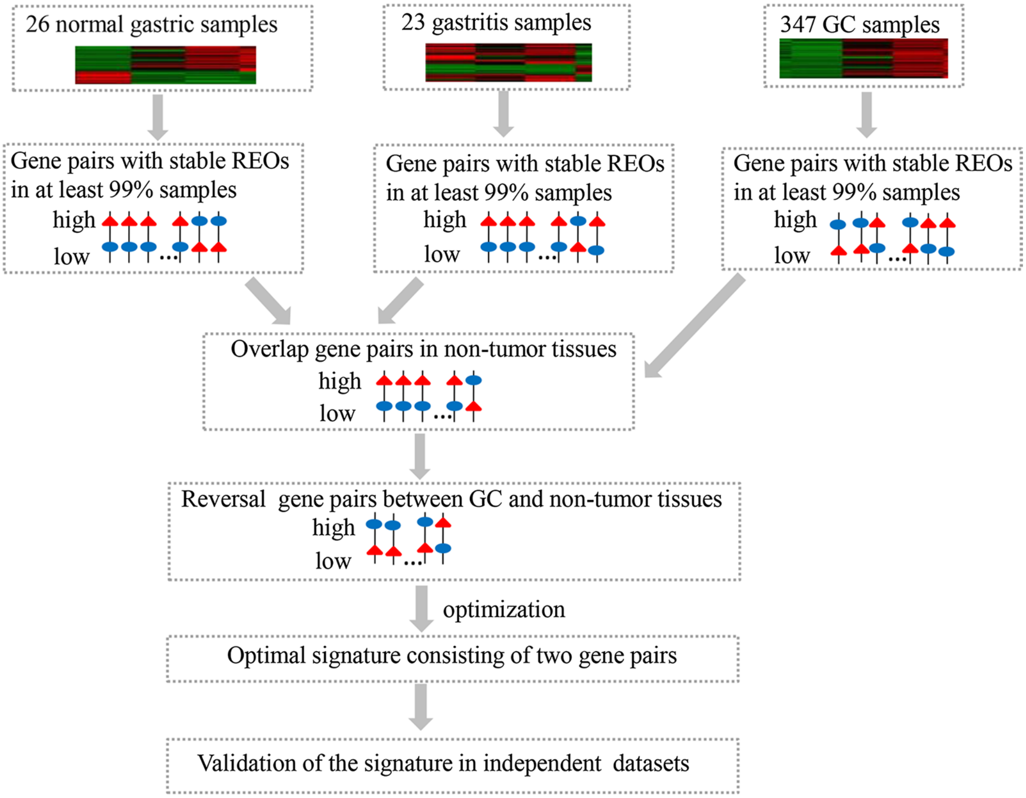
26 normal gastric samples
23 gastritis samples
347 GC samples
Gene pairs with stable REOs in at least 99% samples
high
low
Gene pairs with stable REOs in at least 99% samples
high
low
Gene pairs with stable REOs in at least 99% samples
high
low
Overlap gene pairs in non-tumor tissues
high
low
Reversal gene pairs between GC and non-tumor tissues
high
low
optimization
Optimal signature consisting of two gene pairs
Validation of the signature in independent datasets

Fig. 1 Outline of the processes for developing and validating the GC diagnosis signature

training data (see Methods). According to the reversal degrees of the six gene pairs, we took the top *k* (1, 2,..., 6) gene pairs as a signature and calculated its classification accuracy (Fig. 2). Finally, the top two gene pairs consisting of three genes, were defined as the diagnosis signature (Table 3). In the training data, all the 26 gastric normal and 23 gastritis tissues were correctly classified as non-GC samples, and all the 347 GC tissues were correctly classified as cancer samples. The AUC and the accuracy were 0.99 and 100%, respectively. The detailed classification accuracy of the signature in each of the training datasets was shown in Additional file 5: Table S3.

Validating the signature

The gene expression profiles of gastric tissues sampled by gastroscopic biopsy or surgical resection were used to validate the performance of the qualitative signature.

Non-GC tissues, including normal, gastritis adjacent-normal, gastritis and intestinal metaplasia tissues, from non-GC patients were all sampled by gastroscopic biopsy. The result showed that 96.20% of the 79 non-GC tissues from GSE5081, GSE60662, GSE106656 and GSE34619 were correctly identified as non-GC (Table 4 and Additional file 6). For gastroscopic biopsy specimens, 96.84% of the 158 GC tissues from the GSE14210 and GSE52138 datasets and six of seven GC adjacent-normal tissues from the GSE52138 dataset were correctly identified as GC (Table 4 and Additional file 6). For surgical resection specimens, as described in Table 2, 98.37% of 2560 GC tissues and 97.28% of 221 samples were correctly identified as GC (Table 4). The surgical resection specimens were measured by multiple platforms including the Affymetrix, Illumina and RNA-seq platforms. For the Affymetrix and Illumina platforms used in training data, 99.77% of the 2185 GC tissues and all the 189 GC adjacent-normal tissues were correctly classified to GC tissues. Moreover, 95.73% of the 375 GC tissues and 81.25% of the 32 GC adjacent-normal tissues measured by RNA-seq were correctly classified to GC given that no RNA-seq data participated in training the signature. Especially, 97.67% of the 257 GC patients at stage I were correctly identified as GC. The accuracy and AUC of the validation data were 98.55% and 0.99 (95% CI = 0.95–1, Fig. 3).

To further validate the signature, using RNA-seq platform, we additionally measured gene expression profiles of 21 GC tissues from seven different GC patients, each with three specimens sampled from three tumor locations with different proportions of the tumor epithelial cell (see Table 1). All the 21 GC tissues were correctly

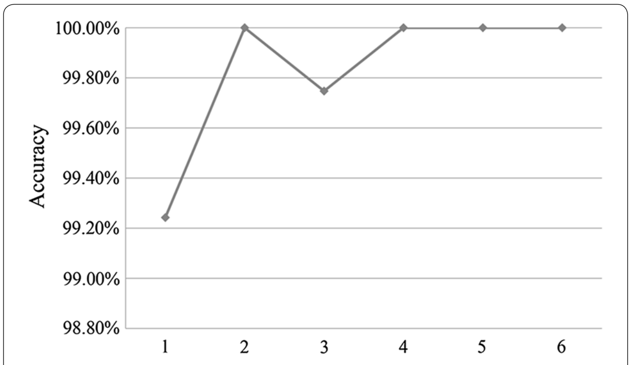

Fig. 2 The classification accuracy of the top k gene pairs in the training data

Table 3 The signature of gene pairs for GC diagnosis

Gene pairs	Gene A	Gene B
Pair 1	CYR61	MMP28
Pair 2	CYR61	ACOX1

The expression level of Gene A is higher than that of Gene B in GC patients

Table 4 The performance of the signature in each of the validation datasets

Platforms	Dataset	Number (sensitivity) of GC tissues	Number (specificity) of non-GC tissues
Affymetrix	GSE5081[a]	–	32 (100.00%)
	GSE52138[a]	13 (92.31%)	–
	GSE14210[a]	145 (97.24%)	–
	GSE106656[a]	–	21 (90.48%)
	GSE34619[a]	–	10 (100.00%)
	GSE34942	56 (100.00%)	–
	GSE22377	43 (100.00%)	–
	GSE29272	134 (100.00%)	–
	GSE19826	12 (100.00%)	–
	GSE35809	70 (100.00%)	–
	GSE51105	94 (100.00%)	–
	GSE15459	200 (100.00%)	–
	GSE62254	300 (99.00%)	–
Illumina	GSE13861	65 (100.00%)	–
	GSE38024	48 (97.92%)	–
	GSE26899	96 (100.00%)	–
	GSE26253	432 (98.84%)	–
	GSE84437	433 (98.15%)	–
	GSE26942	202 (100.00%)	–
Agilent	GSE60662[a]	–	16 (93.75%)
RNA-seq	TCGA	375 (95.73%)	–
	Our-data	21 (100.00%)	–

[a] Denotes the samples collected by gastroscopic biopsy

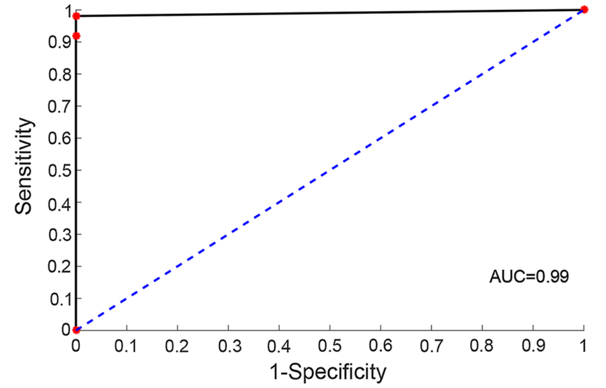

Fig. 3 The receiver characteristic operating curves

classified to GC by our signature, even when the proportion of the tumor epithelial cell was as low as 14% (Table 4).

Together, the above results validated that the signature can accurately discriminate GC, including GC adjacent-normal tissues, from non-GC patients, even when the sampling location is inaccurate.

Discussion

At present, the histological analysis of the gastroscopic biopsy specimen is affected by the sampling location and tissue amount [8]. In this study, a robust qualitative transcriptional signature, including two gene pairs consisting of three genes, was developed to aid the early diagnosis of GC using either gastroscopic biopsy or surgical resection specimens. The signature can accurately distinguish GC tissues from non-GC tissues including normal, gastritis and intestinal metaplasia tissues. As shown in this study, the signature can accurately classify GC tissues to GC when the proportion of the tumor epithelial cell was as low as 14%. Especially, it can identify most of GC adjacent-normal tissues as cancer, suggesting that the signature can identify GC even when the sampling location is inaccurate. Notably, all the non-GC tissues sampled by gastroscopic biopsy can be correctly identified as non-GC. However, the specimens sampled by gastroscopic biopsy for gastritis and intestinal metaplasia are limited, and it deserves further studies using large collections of non-GC specimens.

The amount of the gastroscopic biopsy specimens used in the study was about 1–8 μg total RNA [41–43] which was relatively large. In clinical practice, it is often difficult to obtain sufficient amount of biopsy specimens for gene expression profiling or other molecular measurements [11, 44]. Fortunately, we have shown that the REO-based signatures can be robustly applied to specimens with RNA amplification from as low as 150–250 pg total RNA of cancer cells [31]. Therefore, it is highly possible that the two gene pairs could be used to gastroscopic biopsy specimens with minimum sampling amounts. We compared the expression levels of the two genes in each of the signature gene pairs. The fold changes (FC) of the two genes in each of the signature gene pairs across different datasets for the GC, GC adjacent-normal and non-GC groups were quite different (Additional files 7 and 8). For the gene pair of CYR61 and MMP28, the median values of FC between CYR61 and MMP28 ranged from 1.17 to 30.56 in the GC group across different datasets, while in the non-GC group the median values of FC ranged from 0.76 to 0.89 (Additional file 7: Table S4). Similar results for the gene pair of CYR61 and ACOX1 were also observed (Additional files 7 and 8). Notably, two genes with high expression levels in a sample can hardly reach large FC even if the absolute expression level difference between the two genes is rather large. Besides, two genes with low expression levels in a sample may reach large FC simply due to large measurement variations [45]. To more clearly show the quantitative expression level difference of two genes in each of the signature gene pairs, we also calculated the value of the expression level of CYR61 minus the expression level of MMP28 (ACOX1) in a sample as a measure to show the difference of the two genes consisting of the signature gene pairs (Additional files 9 and 10**).** The median values of the subtraction of MMP28 from CYR61 ranged from 1.30 to 1868.50 in the GC group across different datasets, while in the non-GC group the median values ranged from −2.29 to −0.73 (Additional file 9: Table S5). The results were similar for the gene pair of CYR61 and ACOX1 (Additional files 9 and 10). The subtraction values were quite different for different platforms. However, they varied even in the same platform. For example, the median values of the subtraction of MMP28 from CYR61 in GC group ranged from 2.84 to 1868.5 for GPL6947 (Additional files 9 and 10). The above results showed that the subtle quantitative difference (such as FC and subtraction) of each of the signature gene pairs is quite different across different samples for both the GC and non-GC groups because the quantitative gene expression measurements are affected by the measurement batch effects and many other factors such like the sample quality [29, 31, 46]. However, the REOs of the gene pairs in each group are very stable.

We additionally evaluated the performance of the signature on other types of cancers including liver, colorectal and pancreatic cancers (Additional file 11: Table S6). As shown in Additional file 12: Table S7, the results showed that the signature was unsuitable for these types of cancers. Notably, the signature can classify cancer tissues of liver, colorectum and pancreas as cancer although it cannot correctly classify most non-cancer tissues as non-cancer. The signature genes, including CYR61, MMP28 and ACOX1, may play important roles in the initiation and progression of cancer. As shown in Additional file 13: Table S8, CYR61 and MMP28 are involved in functions such as cell proliferation, differentiation or metastasis related to the initiation and progression of cancer. ACOX1 has been reported to regulate cancer development [47] and its dysfunction is linked to hepatocarcinogenesis [48] and migration and invasion of colorectal cancer cells [49]. Therefore, the stable REOs of genes in the signature may be an inherent feature of cancer which deserves our future study.

Conclusions

In summary, we have developed a transcriptional qualitative signature for GC diagnosis, which exhibits robust and excellent performance in data measured by different laboratories with different platforms.

Additional files

Additional file 1: Table S1. The baseline characteristics of seven GC patients.

Additional file 2. The code to identify the signature for GC diagnosis. All of the analysis programs to develop the diagnostic signature were written using the R language (R 3.1.3).

Additional file 3: Table S2. The number of stable and reversal gene pairs identified in the training data.

Additional file 4. The REOs of the top gene pairs. The distributions of REOs of the top six gene pairs in each of the training datasets.

Additional file 5: Table S3. The classification accuracy of the signature in each of the training datasets.

Additional file 6. The REOs of the signature gene pairs. The distributions of REOs of the signature gene pairs in each of the validation datasets.

Additional file 7: Table S4. The median values of FC of each signature gene pair across different datasets for the GC, non-GC and GC adjacent-normal groups.

Additional file 8: Fig. S1. The distributions of FCs of each signature gene pairs across different datasets for the GC, non-GC and GC adjacent-normal groups. Gene pair1 and gene pair2 represent gene pairs of CYR61-MMP28 and CYR61-ACOX1, respectively.

Additional file 9: Table S5. The median values of the subtraction of two gene expression levels across different datasets for the GC, non-GC and GC adjacent-normal groups.

Additional file 10: Fig. S2. The distributions of the subtraction of two gene expression levels across different datasets for the GC, non-GC, and GC adjacent-normal groups.

Additional file 11: Table S6. The datasets of cancer and non-cancer tissues for liver, colorectum and pancreas.

Additional file 12: Table S7. The performance of the signature in classifying cancer and non-cancer tissues of liver, colorectum and pancreas.

Additional file 13: Table S8. The summary of genes in the signature.

Abbreviations

GC: gastric cancer; REOs: relative expression orderings; GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas; AUC: area under curve; ROC: receiver operating characteristic; CI: confidence intervals; FC: fold changes.

Authors' contributions

ZG, CHZ and HDY conceived and designed the overall study. MFL, HFC and BTZ performed the computational experiments. QZG, MFL, HXC, LLC and WBZ participated in study design. HDY and ZG wrote the manuscript. ZG, HDY, MFL, HFC and HML revised the manuscript. All authors read and approved the final manuscript.

Author details

[1] Department of Bioinformatics, Key Laboratory of Ministry of Education for Gastrointestinal Cancer, School of Basic Medical Sciences, Fujian Medical University, Fuzhou 350122, China. [2] Department of Gastric Surgery, Fujian Medical University Union Hospital, No. 29 Xinquan Road, Fuzhou 350001, China. [3] Department of General Surgery, Fuzhou Second Hospital Affiliated To Xiamen University, Xiamen 350007, China. [4] Department of Systems Biology, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, China.

Acknowledgements

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The gene expression profiles of 21 GC specimens from seven GC patients are available at the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/) under Accession Number GSE116782.

Consent for publication

All the authors in this paper consent to publication of the work.

Ethics approval and consent to participate

This study was approved by the institutional review boards of all participating institutions, and written consent forms were obtained from all participants.

Funding

This work was supported by the National Natural Science Foundation of China [Grant Numbers 61801118, 81372213, 81501215, 81501829 and 81402738] and Education research project for young and middle-aged teachers in Fujian province [Grant Number JAT170214]. Startup Fund for scientific research, Fujian Medical University [Grant Number 2017XQ2002].

Publisher's Note

Received: 26 September 2018 Accepted: 21 February 2019
Published online: 28 February 2019

References

1. Yang L, Parkin DM, Ferlay J, Li L, Chen Y. Estimates of cancer incidence in China for 2000 and projections for 2005. Cancer Epidemiol Biomarkers Prev. 2005;14:243–50.
2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA Cancer J Clin. 2013;63:11–30.
3. Chen WQ, Zhang SW, Zou XN, Zhao P. Cancer incidence and mortality in china, 2006. Chin J Cancer Res. 2011;23:3–9.
4. Kim JP, Hur YS, Yang HK. Lymph node metastasis as a significant prognostic factor in early gastric cancer: analysis of 1,136 early gastric cancers. Ann Surg Oncol. 1995;2:308–13.

5. Shiozawa N, Kodama M, Chida T, Arakawa A, Tur GE, Koyama K. Recurrent death among early gastric cancer patients: 20-years' experience. Hepatogastroenterology. 1994;41:244–7.
6. Tan YK, Fielding JW. Early diagnosis of early gastric cancer. Eur J Gastroenterol Hepatol. 2006;18:821–9.
7. Lustosa SA, Saconato H, Atallah AN, Filho GJ, Matos D. Impact of extended lymphadenectomy on morbidity, mortality, recurrence and 5-year survival after gastrectomy for cancer. Meta-analysis of randomized clinical trials. Acta Cir Bras. 2008;23:520–30.
8. Bhandari S, Shim CS, Kim JH, Jung IS, Cho JY, Lee JS, Lee MS, Kim BS. Usefulness of three-dimensional, multidetector row CT (virtual gastroscopy and multiplanar reconstruction) in the evaluation of gastric cancer: a comparison with conventional endoscopy, EUS, and histopathology. Gastrointest Endosc. 2004;59:619–26.
9. Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. Cancer Epidemiol Biomarkers Prev. 2014;23:700–13.
10. Tajiri H, Ohtsu A, Boku N, Muto M, Chin K, Matsumoto S, Yoshida S. Routine endoscopy using electronic endoscopes for gastric cancer diagnosis: retrospective study of inconsistencies between endoscopic and biopsy diagnoses. Cancer Detect Prev. 2001;25:166–73.
11. Kim YJ, Park JC, Kim JH, Shin SK, Lee SK, Lee YC, Chung JB. Histologic diagnosis based on forceps biopsy is not adequate for determining endoscopic treatment of gastric adenomatous lesions. Endoscopy. 2010;42:620–6.
12. Park JY, von Karsa L, Herrero R. Prevention strategies for gastric cancer: a global perspective. Clin Endosc. 2014;47:478–89.
13. Itabashi M, Hirota T, Unakami M, Ueno M, Oguro Y, Yamada T, Kitaoka H, Ichikawa H. The role of the biopsy in diagnosis of early gastric cancer. Jpn J Clin Oncol. 1984;14:253–70.
14. Suvakovic Z, Bramble MG, Jones R, Wilson C, Idle N, Ryott J. Improving the detection rate of early gastric cancer requires more than open access gastroscopy: a five year study. Gut. 1997;41:308–13.
15. Amin A, Gilmour H, Graham L, Paterson-Brown S, Terrace J, Crofts TJ. Gastric adenocarcinoma missed at endoscopy. J R Coll Surg Edinb. 2002;47:681–4.
16. Yalamarthi S, Witherspoon P, McCole D, Auld CD. Missed diagnoses in patients with upper gastrointestinal cancers. Endoscopy. 2004;36:874–9.
17. Voutilainen ME, Juhola MT. Evaluation of the diagnostic accuracy of gastroscopy to detect gastric tumours: clinicopathological features and prognosis of patients with gastric cancer missed on endoscopy. Eur J Gastroenterol Hepatol. 2005;17:1345–9.
18. Vradelis S, Maynard N, Warren BF, Keshav S, Travis SP. Quality control in upper gastrointestinal endoscopy: detection rates of gastric cancer in Oxford 2005–2008. Postgrad Med J. 2011;87:335–9.
19. Hosokawa O, Tsuda S, Kidani E, Watanabe K, Tanigawa Y, Shirasaki S, Hayashi H, Hinoshita T. Diagnosis of gastric cancer up to three years after negative upper gastrointestinal endoscopy. Endoscopy. 1998;30:669–74.
20. Nakajima T, Maekita T, Oda I, Gotoda T, Yamamoto S, Umemura S, Ichinose M, Sugimura T, Ushijima T, Saito D. Higher methylation levels in gastric mucosae significantly correlate with higher risk of gastric cancers. Cancer Epidemiol Biomarkers Prev. 2006;15:2317–21.
21. Baba Y, Ishimoto T, Kurashige J, Iwatsuki M, Sakamoto Y, Yoshida N, Watanabe M, Baba H. Epigenetic field cancerization in gastrointestinal cancers. Cancer Lett. 2016;375:360–6.
22. Yap YL, Zhang XW, Smith D, Soong R, Hill J. Molecular gene expression signature patterns for gastric cancer diagnosis. Comput Biol Chem. 2007;31:275–87.
23. Fan ZY, Liu W, Yan C, Zhu ZL, Xu W, Li JF, Su L, Li C, Zhu ZG, Liu B, Yan M. Identification of a five-lncRNA signature for the diagnosis and prognosis of gastric cancer. Tumour Biol. 2016;37:13265–77.
24. Chen S, Li T, Zhao Q, Xiao B, Guo J. Using circular RNA hsa_circ_0000190 as a new biomarker in the diagnosis of gastric cancer. Clin Chim Acta. 2017;466:167–71.
25. Liu R, Zhang C, Hu Z, Li G, Wang C, Yang C, Huang D, Chen X, Zhang H, Zhuang R, et al. A five-microRNA signature identified from genome-wide serum microRNA expression profiling serves as a fingerprint for gastric cancer diagnosis. Eur J Cancer. 2011;47:784–91.
26. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8:118–27.
27. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet. 2010;11:733–9.
28. Nygaard V, Rodland EA, Hovig E. Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. Biostatistics. 2016;17:29–39.
29. Guan Q, Yan H, Chen Y, Zheng B, Cai H, He J, Song K, Guo Y, Ao L, Liu H, et al. Quantitative or qualitative transcriptional diagnostic signatures? A case study for colorectal cancer. BMC Genomics. 2018;19:99.
30. Chen X, Guo X, He P, Nie J, Yan X, Zhu J, Zhang L, Mao G, Wu H, Liu Z, et al. Interactive influence of N6AMT1 and As3MT genetic variations on arsenic metabolism in the population of inner mongolia, China. Toxicol Sci. 2017;155:124–34.
31. Liu H, Li Y, He J, Guan Q, Chen R, Yan H, Zheng W, Song K, Cai H, Guo Y, et al. Robust transcriptional signatures for low-input RNA samples based on relative expression orderings. BMC Genomics. 2017;18:913.
32. 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;70:10202–12.
33. Villanueva A, Portela A, Sayols S, Battiston C, Hoshida Y, Mendez-Gonzalez J, Imbeaud S, Letouze E, Hernandez-Gea V, Cornella H, et al. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. Hepatology. 2015;61:1945–56.
34. Zheng G, Xiong Y, Xu W, Wang Y, Chen F, Wang Z, Yan Z. A two-microRNA signature as a potential biomarker for early gastric cancer. Oncol Lett. 2014;7:679–84.
35. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
36. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12:357–60.
37. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5.
38. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4:249–64.
39. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29–36.
40. Ao L, Zhang Z, Guan Q, Guo Y, Guo Y, Zhang J, Lv X, Huang H, Zhang H, Wang X, Guo Z. A qualitative signature for early diagnosis of hepatocellular carcinoma based on relative expression orderings. Liver Int. 2018;38:1812–9.
41. Kim HK, Choi IJ, Kim HS, Kim JH, Kim E, Park IS, Chun JH, Kim IH, Kim IJ, Kang HC, et al. DNA microarray analysis of the correlation between gene expression patterns and acquired resistance to 5-FU/cisplatin in gastric cancer. Biochem Biophys Res Commun. 2004;316:781–9.
42. Galamb O, Gyorffy B, Sipos F, Dinya E, Krenacs T, Berczi L, Szoke D, Spisak S, Solymosi N, Nemeth AM, et al. Helicobacter pylori and antrum erosion-specific gene expression patterns: the discriminative role of CXCL13 and VCAM1 transcripts. Helicobacter. 2008;13:112–26.
43. Kim HK, Choi IJ, Kim CG, Kim HS, Oshima A, Michalowski A, Green JE. A gene expression signature of acquired chemoresistance to cisplatin and fluorouracil combination chemotherapy in gastric cancer patients. PLoS ONE. 2011;6:e16694.
44. Lee CK, Chung IK, Lee SH, Kim SP, Lee SH, Lee TH, Kim HS, Park SH, Kim SJ, Lee JH, et al. Is endoscopic forceps biopsy enough for a definitive diagnosis of gastric epithelial neoplasia? J Gastroenterol Hepatol. 2010;25:1507–13.
45. Ao L, Yan H, Zheng T, Wang H, Tong M, Guan Q, Li X, Cai H, Li M, Guo Z. Identification of reproducible drug-resistance-related dysregulated genes in small-scale cancer cell line experiments. Sci Rep. 2015;5:11895.
46. Chen R, Guan Q, Cheng J, He J, Liu H, Cai H, Hong G, Zhang J, Li N, Ao L, Guo Z. Robust transcriptional tumor signatures applicable to both formalin-fixed paraffin-embedded and fresh-frozen samples. Oncotarget. 2017;8:6652–62.

47. Huang J, Viswakarma N, Yu S, Jia Y, Bai L, Vluggens A, Cherkaoui-Malki M, Khan M, Singh I, Yang G, et al. Progressive endoplasmic reticulum stress contributes to hepatocarcinogenesis in fatty acyl-CoA oxidase 1-deficient mice. Am J Pathol. 2011;179:703–13.
48. Chen XF, Tian MX, Sun RQ, Zhang ML, Zhou LS, Jin L, Chen LL, Zhou WJ, Duan KL, Chen YJ, et al. SIRT5 inhibits peroxisomal ACOX1 to prevent oxidative damage and is downregulated in liver cancer. EMBO Rep. 2018;19:e45124.
49. Sun LN, Zhi Z, Chen LY, Zhou Q, Li XM, Gan WJ, Chen S, Yang M, Liu Y, Shen T, et al. SIRT1 suppresses colorectal cancer metastasis by transcriptional repression of miR-15b-5p. Cancer Lett. 2017;409:104–15.