Contents lists available at ScienceDirect

EBioMedicine

Research paper

Clinically applicable 53-Gene prognostic assay predicts chemotherapy benefit in gastric cancer: A multicenter study

Linghua Zhu[a,$], Haifeng Wang[b,$], Chengfei Jiang[c,$], Wenhuan Li[d], Shuting Zhai[a], Xiujun Cai[a], Xianfa Wang[a], Linghong Liao[e], Feng Tao[b], Dexi Jin[d], Guofu Chen[d], Yankai Xia[f], Jian-Hua Mao[g], Bin Li[h,*], Pin Wang[c,*], Bo Hang[g,*]

[a] Department of General Surgery, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China
[b] Department of Gastrointestinal Surgery, Shaoxing People's Hospital (Shaoxing Hospital of Zhejiang University), Shaoxing, Zhejiang, China
[c] Department of Gastroenterology, Nanjing Drum Tower Hospital, Affiliated Hospital of Nanjing University Medical School, Nanjing, Jiangsu, China
[d] Department of Gastrointestinal Surgery, The First People's Hospital of Wenling, Wenling, Zhejiang, China
[e] Fujian Key Laboratory of TCM Health State, Fujian University of Traditional Chinese Medicine, Fuzhou, Fujian, China
[f] School of Public Health, Nanjing Medical University, Nanjing, Jiangsu, China
[g] Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States
[h] Nanjing KDRB Biotech Inc., Ltd, Jiangning District, Nanjing, Jiangsu, China

ARTICLE INFO

ABSTRACT

Background: We previously established a 53-gene prognostic signature for overall survival (OS) of gastric cancer patients. This retrospective multi-center study aimed to develop a clinically applicable gene expression detection assay and to investigate the prognostic value of this signature.

Methods: A TCGA gastric adenocarcinoma cohort (TCGA-STAD) was used for comparing 53-gene signature with other gene signatures. A high-throughput mRNA hybridization gene expression assay was developed to quantify the expression of 53-genes in formalin-fixed paraffin-embedded tissues of 540 patients enrolled from three hospitals. 180 patents were randomly selected from two hospitals to build a prognostic prediction model based on the 53-gene signature using leave-p-out (one-third out) cross-validation method together with Cox regression and Kaplan-Meier analysis, and the model was assessed on three validation cohorts.

Findings: In the evaluation phase, studies based on TCGA-STAD showed that the 53-gene signature was significantly superior to other three prognostic signatures and was independent of TCGA molecular subtypes and clinical factors. For clinical validation and utility, the prognostic scores were generated using the newly developed assay, which was reliable and sensitive, in 100 sampling training sets and were significantly associated with OS in 100 sampling validation sets. The scores were significantly associated with OS in three independent and combined validation cohorts, and in patients with stages II and III/IV. The multivariate Cox regression demonstrated that the prognostic power of the score was independent of clinical factors, consistent with those findings in the TCGA dataset. Finally, patients with good prognostic scores exhibited significantly a better 5-year OS rate from adjuvant FOLFOX chemotherapy after surgery than from other chemotherapies.

Interpretation: The 53-gene prognostic score system is clinically applicable for predicting the OS of patients independent of clinical factors in gastric cancers, which could also be a promising predictive biomarker for FOLFOX regimen.

Funding: Chinese National Science and Technology, National Natural Science Foundation and Natural Science Foundation of Jiangsu Province.

* Corresponding authors.
E-mail addresses: libin1@gmail.com (B. Li), pinwang729@126.com (P. Wang), Bo_hang@lbl.gov (B. Hang).
$ Linghua Zhu, Haifeng Wang and Chengfei Jiang share co-first authorship.

Research in Context

Evidence before this study

The 5-year overall survival of patients with gastric cancer remains generally poor. Previously, we employed bioinformatics methods to identify a novel 53-gene prognostic signature for gastric cancer patients, which robustly predicts their overall survival. Later, several reports have analyzed gene expression patterns to predict patient outcomes such as survival and benefit from adjuvant therapy, but no multigene prognostic biomarkers have yet been used or recommended for clinical utility.

Added value of this study

We first used a TCGA-STAD cohort to demonstrate that our 53-gene signature was more robust than three published prognostic signatures in predicting overall survival and is independent of molecular subtypes and clinical factors that are associated with patient outcomes. To deploy the clinical application of the 53-gene prognostic signature, we first developed a reliable high-throughput mRNA hybridization-based assay, which provides a solution to better measure the 53-gene expression in formalin-fixed paraffin-embedded (FFPE) tissues from three hospitals. Using such a clinically applicable method and the cross-validation approach, we built a prognostic prediction model based on the data of the 53-gene signature in gastric cancer patients. The significant predictive power of the score system in patients with II to IV stages of gastric cancer was successfully validated in three independent hospital cohorts. The score system was also proven to be independent of clinical factors. Finally, we observed that patients with good scores can significantly benefit from FOLFOX chemotherapy after gastrectomy.

Implications of all the available evidence

Our data indicate that the 53-gene prognostic score is clinically applicable for predicting the overall survival of gastric cancer patients independent of clinical factors. It could also be a promising predictive biomarker for FOLFOX regimen in patients with resectable gastric cancer. Future work should focus on prospective clinical trials to fully deploy this score system into clinical utility.

1. Introduction

Gastric (stomach) cancer currently ranks as the fifth most common cancer and the third leading cause of cancer-related death around the world (1.2). East Asia has about 50% cases of this disease, with China being the most affected nation (accounting for 42.6% of the global incidence and 45% of all gastric cancer-related deaths) [3]. Curative surgery in combination with adjuvant chemotherapy is the current most common treatment plan for stage II-III gastric cancer. The 5-year overall survival remains generally poor [4], which may be contributed by many factors including clinical, histopathological and genetic differences [5]. Hence, it is quite challenging to identify those factors that are critical and independent for predicting patient clinical outcome to provide a more accurate risk assessment for personalized treatment.

In recent years, microarray technology and next generation sequencing (NGS)/genome wide association studies (GWAS) have become invaluable tools to deconvolute the heterogeneity of gastric cancer and genetic susceptibility to the disease [6–8]. For instance, The Cancer Genome Atlas (TCGA) has identified four molecular

subtypes in gastric cancer through comprehensive molecular profiling, including Epstein-Barr virus (EBV)-positive, microsatellite unstable (MSI), genomically stable (GS) and chromosomal instability (CIN) tumors [9]. Such classification reflects both background genetics and molecular pathogenetic features. Gene expression patterns have also recently been used as prognostic biomarkers in various types of cancer [*e.g.*, [refs 10-16]]. The power of such analysis has been well demonstrated with both Oncotype DX and MammaPrint assays in predicting clinical outcome of patients with breast cancer [10, 11]. In gastric cancer, gene expression patterns have been analyzed to predict patient outcomes such as recurrence, metastasis, and benefit from adjuvant chemotherapy [17–24], but, to our knowledge, extended validation of bioinformatics findings is rare with prognostic biomarker signatures, and no multigene biomarkers have yet been recommended for guiding treatment plans. Previously, we used public genomic datasets from TCGA and NCBI Gene Expression Omnibus (NCBI-GEO) to establish a 53-gene prognostic score system that strongly predicted gastric cancer patients available in independent online databases to have either a poor or good overall survival [8]. In this study, using a TCGA gastric adenocarcinoma cohort, we further demonstrated that this signature is independent of the molecular subtypes described above and is superior in prognostic power in comparison to three other prognostic multigene signatures [22-24].

For clinical validation and application of such a multigene signature, it is essential to have a reliable, sensitive, and high-throughput assay for measuring gene expression using suitable patient samples. In the past years, RT-qPCR quantifies mRNA levels of prognostic genes for such a purpose. This is best demonstrated by the 21-gene Oncotype DX assays developed by Genomic Health [10, 13], as RT-PCR is a mature technology routinely used in predictive assays in clinical settings. Here, we developed a modified hybridization-based assay for the quantitative measurement of mRNA in routinely prepared formalin fixed paraffin-embedded (FFPE) tumor blocks. This assay provides a 96-well high-throughput setting and could be more reliable than RT-PCR to detect RNA or DNA signal in archived FFPE samples [23, 25, 26]. With this assay, we successfully validated the prognostic power of the 53-gene score on the OS of 540 patients with stage I to IV gastric cancer. To the best of our knowledge, this is the first retrospective clinical study using the hybridization-based assay for validating a gene-expression signature in gastric cancer patients with a multi-center, relatively large-sized patient cohort. Finally, we showed that our score system predicts adjuvant chemotherapy benefit in gastric cancer, when the effect of FOLFOX (leucovorin, fluorouracil, and oxaliplatin) chemotherapy and other chemotherapies were analyzed for patient survival in different prognostic score groups.

2. Methods

The multi-step strategy employed in this work included further evaluation of the prognostic significance of the 53-gene signature using TCGA data and clinical validation and utility of this score system using specimens from three hospitals.

2.1. TCGA data-based gene expression analysis

The mRNA expression levels in the TCGA gastric adenocarcinoma (TCGA-STAD) dataset (Firehose Legacy) were downloaded from cBio-Portal (https://www.cbioportal.org), which contains 407 specimens with survival information. The bioinformatics approach used to compare the performance of our 53-gene signature with three published prognostic signatures, which contain 4, 6 and 23 genes, respectively [22-24] was summarized in Supplementary Figure 1. We performed a multivariate Cox regression analysis on 100 training sets (271 patients), averaged the coefficient values and calculated prognostic score for each patient. The patients in each signature were divided

into tertiles in each resampling training set based on their scores and the scores at the cut-points were recorded. The Kaplan-Meier (K-M) plotter and Cox regression analysis were used to define the differences of OS among the three risk groups in 100 resampling validation sets. Finally, hazard ratio (HR) values were calculated for each testing set for the "intermediate" and "poor" groups in comparison to the "good" group.

We also analyzed the possible enrichment of our 53-gene signature in the four molecular subtypes of gastric cancer patients (369 specimens eligible). The K-M analysis, including OS, log-rank p-value and HR with 95% confidence intervals (CI), within the four molecular subtypes was performed and the percent distribution of the patients in three prognostic scores for each of the molecular subtypes was recorded.

Finally, we comprehensively analyzed the HR and its 95% CI values of various clinical factors using univariate Cox regression analysis in TCGA-STAD, including age, gender, TNM staging, Lauren classification, WHO classification, primary tumor site and lymph node ratio, as well as four molecular subtypes, to screen those with significant impact on patient survival. We then selected such parameters for multivariate Cox regression analysis, in relation to the 53-gene prognostic score.

2.2. Multi-hospital validation of prognostic significance of the 53-Gene expression assay

2.2.1. Study design, cohorts and tissue specimens

The strategy of the study with our clinical data was two-fold. The first was to develop a nucleic acid hybridization-based assay for clinically applicable measurement of the expression levels of 53 genes plus 5 reference genes in gastric cancer. The second was to build a prognostic prediction model on 5-year OS based on the 53-gene signature and then validate the model using independent patient cohorts (Fig. 3). The 5-year OS was defined as the time from the first pathological diagnosis to death from any cause. A total of 540 patients were enrolled who were administered as pathologically confirmed gastric adenocarcinoma in various stages and treated with gastrectomy between 2008 and 2013 (at least 5-year follow-up), including 192 from Sir Run Run Shaw Hospital, Zhejiang University (Hangzhou, Zhejiang Province), 212 from Nanjing Drum Tower Hospital, Nanjing University (Nanjing, Jiangsu Province), and 136 from Shaoxing People's Hospital, Zhejiang University (Shaoxing, Zhejiang Province). All patients must have complete clinicopathological information including age at diagnosis, sex, histopathologic staging, treatment regimens (gastrectomy and adjuvant chemotherapy) and survival status. Follow-up visits by the patients or telephone calls by the medical staff were carried out till October 2019 to ensure accurate and updated 5-year survival information.

Totally, 506 patients were treated with various regimens and/or combinations of adjacent chemotherapy after gastrectomy, of which 121 cases had no definitive information. A total of 250 patients received first-line chemotherapy, including 129 cases underwent FOLFOX regimen (the combination of leucovorin (folinic acid), fluorouracil, and oxaliplatin) and 121 cases with other drugs and combinations (non-FOLFOX).

All patient samples were archived FFPE tissue blocks of gastric adenocarcinoma from 2018 to 2013. Before making the paraffin slices for this study, all FFPE specimen were checked by a pathologist through reviewing hematoxylin and eosin-stained slides to ensure at least 70% of the section area to be gastric cancer tissue. For each patient, 6 slices of 6 μm-thickness unstained tissue block were collected from the Department of Pathology at each hospital and stored in the Eppendorf tubes at $-20\,°C$ until use.

Ethics statement

This study was approved by the Ethics Committees of the three participating hospitals (Sir Run Run Shaw Hospital (document number: Research 20,190,717−3), Nanjing Drum Tower Hospital (2019−196−01), and Shaoxing People's Hospital (2019-K-Y-264−01), and written informed general consent was obtained from each patient.

2.2.2. 53-Gene expression assay and data treatment

FFPE samples were lysed with homogenizing buffer supplemented with 2μl Proteinase K (50 μg/ml) at 65 °C for six hours, followed by a short centrifugation to separate the tissue homogenates from debris and wax. The mRNA level of 53 prognostic genes and 5 reference genes (the gene bank NM numbers are given in Supplementary Table 4) in FFPE samples was determined using Quantigene Plex 2.0 reagent system (Panomics/Affymetix, Inc., Fremont, CA). Individual gene specific probe sets and bead specific capture extenders were designed using ProbeDesigner (Bayer Corp., Emeryville, CA) and optimized by Panomics/Affymetrix, Inc. All oligonucleotides were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA). Assays were performed in a 96-well plate according to manufacturers' protocol with small modifications to optimize the S/N ratio. Briefly, the beads, tissue homogenates and probe set hybridization was carried out at 54 °C overnight. Luminex beads were sequentially hybridized (45 min, 50 °C) with pre-amplifier probe, the amplifier probe, the double-biotin label probe and streptavidin-conjugated *R*-phycoerythrin (SAPE) with washes in between these steps. The resulting fluorescence signal associated with individual capture beads is read on a Luminex-200 system and raw data were collected as median fluorescence intensity (MFI). Net MFI values were utilized for further analysis, which is to subtract blank wells with no RNA from MFI values.

The net MFI was treated as follows before analysis [1]: the expression of each gene was normalized using the mean of five reference genes (*ACTB*, *RPLPO, GUSB, TFRC*, and *GAPDH*) for each patient (see e-component/Data file S1 for normalized expression data of the 53 genes) [2]; to remove batch variations, the same samples were measured in different batches and between different hospitals and their readings were used to adjust such variations [3]; all the data from three hospitals were converted to Z-score together; and [4] match with clinical information on patient survival for further statistical analysis. Before each experiment, pre-testing was carried out using four samples with 4-fold serial dilutions to ensure the loading quantity is well within the linear range of the assay.

2.2.3. Statistics

Cox regression analysis of 53 gene expression was performed on 100 resampling training sets consisted of 120 patients and the coefficient values for each gene were averaged as weights for calculating prognostic score. At each resampling, we generated 180 random numbers between 0 and 180 that were assigned to each patient, then sorted patients based on the assigned random numbers from minimum to maximum, and finally the top 120 patients were designed as a training set and the remaining 60 patients as a testing set. The prognostic score for each patient was calculated using the formula $= b1$ (average coefficient of gene 1) $*$ x1 (expression level of gene 1) $+ b2 * x2 + \ldots + b53 *$ x53. The patients were then ranked by their scores and divided into tertile in each resampling training set. The tertile cutoff values were then averaged across 100 resampling training sets, which were used to divide patients into good, intermediate and poor prognosis groups. The Kaplan-Meier plotter including K-M survival analysis, log-rank p-value and HR with 95% CI were used to define the differences in OS of the gastric cancer patients in 100 resampling validation sets and three independent hospital cohorts or their pooled cohort (see e-component/Data file S2 for clinical and histologic characteristics of patients in these cohorts). The multivariate Cox regression analysis was employed for age, gender, TNM staging, WHO classification and lymph node ratio to analyze the influence of each variable on OS in relation to the 53-gene prognostic score. The chi-square test was used to determine the OS differences for each risk group (good, intermediate or poor) at 60 months between the two different groups of patients underwent FOLFOX regimen versus other drugs/regimes (non-FOLFOX).

Role of the funding source

The sponsors of this study had no role in the research design; the data collection, analysis, and interpretation; the paper writing; and the decision to submit for publication. The corresponding authors had full access to all data obtained from this study and had full responsibility for the decision to submit for publication.

3. Results

3.1. Evaluation of the 53-Gene expression signature using TCGA-STAD dataset

3.1.1. Comparison with existing prognostic signatures

We first compared the performance of our 53-gene signature with three recently published multigene expression signatures [22-24] using cross-validation approach combined with a multivariate Cox regression analysis (Supplementary Figure 1). In the intermediate vs. good group,

the median HR values of the 53-gene signature were 2.23-, 1.63- and 1.40-fold higher than those of the 4-, 6- and 23-gene panels, respectively. In the poor vs. good group, the median HR values were 4.30-, 1.48- and 1.40-fold higher compared to the 4-, 6- and 23-gene panels, respectively. The differences between our signature and any of the other signatures were significant for both the intermediate vs. good and poor vs. good groups ($p<0.0001$ by Mann-Whitney U test) (Fig. 1). These data indicate that the 53-gene score significantly performs better than other signatures in discriminative ability to determine OS of patients with gastric cancer.

3.1.2. Prognostic impact of the 53-gene score is independent of molecular subtypes

To investigate whether the prognostic impact of the 53-gene signature would be enriched in certain molecular subtype(s) of gastric cancer, we carried out K-M survival analysis in the TCGA-STAD cohort whose samples have been already annotated for molecular subtypes. As shown in Fig. 2a, which show percent distribution of three
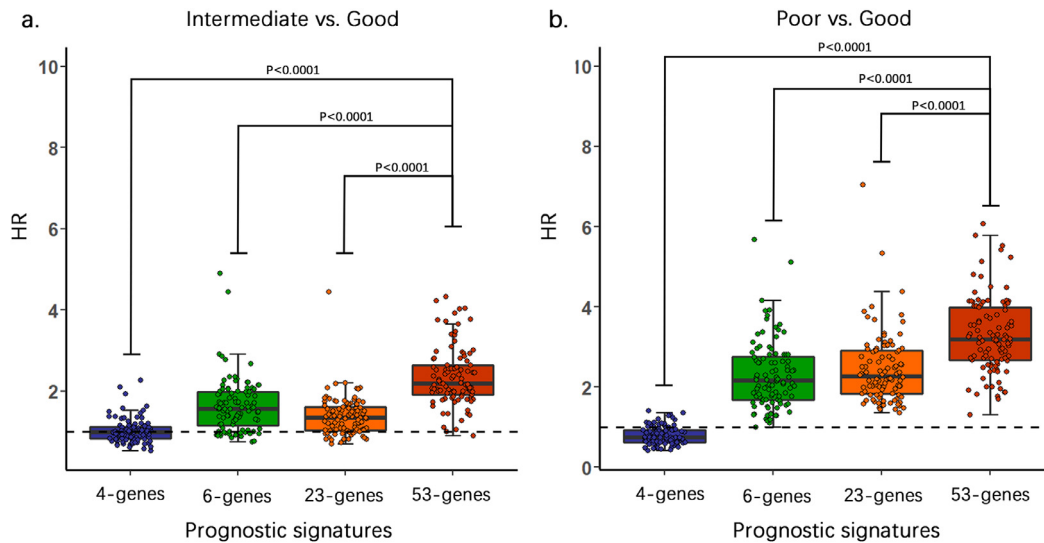

Fig. 1. Comparison of the prognostic performance of the four prognostic signatures in gastric cancer patients. For all signatures, the HR values of all the 100 test sets were calculated using a Cox model based on the prognostic score between groups (intermediate vs. good: left; poor vs. good: right). The differences between the 53-gene signature and other three signatures were significant for both the intermediate vs. good and poor vs. good groups ($p<0.0001$, Mann-Whitney U test).
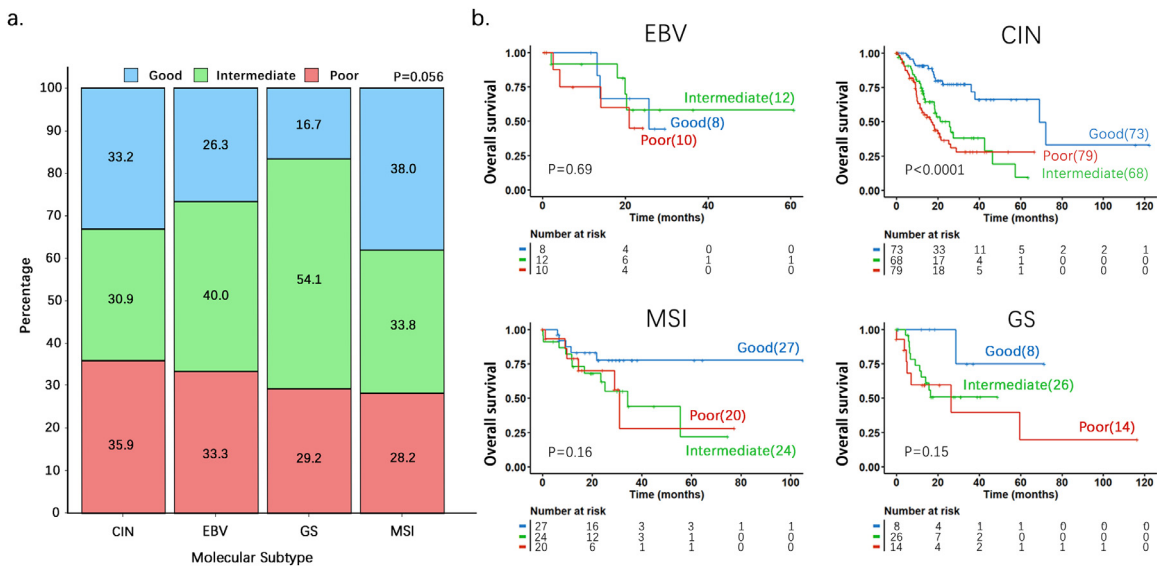

Fig. 2. The Kaplan-Meier analysis on the performance of 53-gene signature in the four molecular subtypes of gastric cancer patients. a. The percent distribution in the molecular subtypes of the patients within the three prognostic score groups. The p-value was obtained by Chi-Square test. b. K-M plots with log-rank p-values were used to determine differences in OS in four molecular subtypes. The p-values were obtained by log-rank test.

Table 1
Multivariate analysis of potential prognostic factors by Cox regression (TCGA-STAD dataset).

Variables	HR	95% CI for HR		P value
		Lower	Upper	
Clinical factors				
Age	1.041	1.021	1.062	<0.0001
Stage				0.178
Stage III vs. Stage I&II	1.175	0.599	2.305	0.640
Stage IV vs. Stage I&II	2.087	0.861	5.057	0.103
T grade				0.315
T3 vs. T1&T2	1.471	0.883	2.450	0.138
T4 vs. T1&T2	1.295	0.715	2.347	0.394
N grade				0.834
N0 vs. N3	1.222	0.507	2.949	0.655
N1 vs. N3	1.300	0.733	2.307	0.370
N2 vs. N3	1.123	0.659	1.912	0.670
M grade				0.850
M1 vs. M0	1.225	0.525	2.858	0.639
Mx vs. M0	1.160	0.452	2.974	0.758
Lymph node ratio	4.076	1.920	8.652	0.0002
53-Gene signature				<0.0001
Intermediate vs. Good	2.546	1.541	4.206	0.0002
Poor vs. Good	3.941	2.379	6.528	<0.0001

prognostic score (good, intermediate and poor) groups in each of the molecular subtypes, there is no significant difference in distribution among the four subtypes (P = 0.056 by Chi-Square test). The K-M plot and log-rank p-value for each molecular subtype was shown in Fig. 2b. It should be noted that although 53-gene signature exhibited the best prognostic performance in the CIN subtype (P < 0.0001 by log-rank test) compared with other subtypes, this subtype contained most of the patients from this cohort (n = 220), while small sample size for MSI (n = 71), GS (n = 48) and EBV (n = 30) subtypes.

3.1.3. Prognostic impact of the 53-gene score is independent of clinical factors

To investigate whether the prognostic impact of the 53-gene signature is independent of clinical factors that could be associated with clinical outcomes, we first performed univariate Cox regression analysis on all available clinical parameters and molecular subtypes in the TCGA-STAD dataset (Supplementary Table 1). We then selected those that demonstrated significant prognostic impact (P < 0.05 by Wald test) for multivariate Cox regression analysis along with the 53-gene signature (Table 1). Although we found that the prognostic scores (good, intermediate and poor) had significant differences in percent distribution in different stages (Supplementary Figure 2). The 53-gene signature was shown to be an independent prognostic factor (Table 1).

3.2. Development and evaluation of a clinically applicable 53-Gene prognostic assay in gastric cancer

3.2.1. Retrospective cohorts of patients from three hospitals and their clinical characterizations

After screening and follow-up, totally 540 patients from the three hospitals were found eligible for this study. The clinical and tumor histologic information of patients is summarized in Supplementary Table 2. The median ages at diagnosis for all three hospitals are very close to about 60 years old. The cancer occurrence was much prominent in men than in women (410 vs. 122). The stage distribution shows that most of the patients were in stage III for all three hospitals (from 48.2% to 61.3%). The tumor histologic types were according to the WHO classifications. 93.7% patients received chemotherapy and the first-line chemotherapy drugs and regimens used in gastric cancer were further summarized in Supplementary Table 3. Of the 250 patients with first-line chemotherapy, more than 50% of them had undergone FOLFOX as adjuvant chemotherapy after surgical resection. No first-line palliative chemotherapy was included.

3.2.2. Development of a 53-gene expression assay using FFPE specimens

The multigene expression assay using a hybridization-based detection on FFPE gastric cancer tissues was developed to provide a practical quantitative tool for the clinical validation and application in individual patient prognosis. The assay was developed partly on QuantiGene Plex reagent system-based quantitation of gene expression [26] in a high-throughput way. For testing with each batch of samples from the same and different hospitals, pre-tests with various concentrations of the FFPE lysates after a 4-fold serial dilution were carried out to determine an optimal dilution condition for the detection. The data showed that the R2 value for the 4-fold serial dilution is perfect ranging from 0.98 to 1.00 for all four samples tested. Thus, the assay demonstrated a very good detection linearity and the sample loading quantity was appropriately within the linearity range (see e-component/Data file S2).

3.2.3. Establishment of the 53-gene prognostic score in patients with gastric cancer

We developed a strategy to establish and validate the 53-gene prognostic score system based on the above assay using FFPE samples from the three hospitals (Fig. 3). We first randomly selected a cohort of 180 patients from Sir Run Run Shaw Hospital and Nanjing Drum Tower Hospital (90 patients/hospital) to build the prognostic prediction model using leave-p-out cross-validation method with 100 times of resampling (Fig. 4a). The Cox regression analysis was performed on all 100 training sets. The coefficient for each gene was averaged
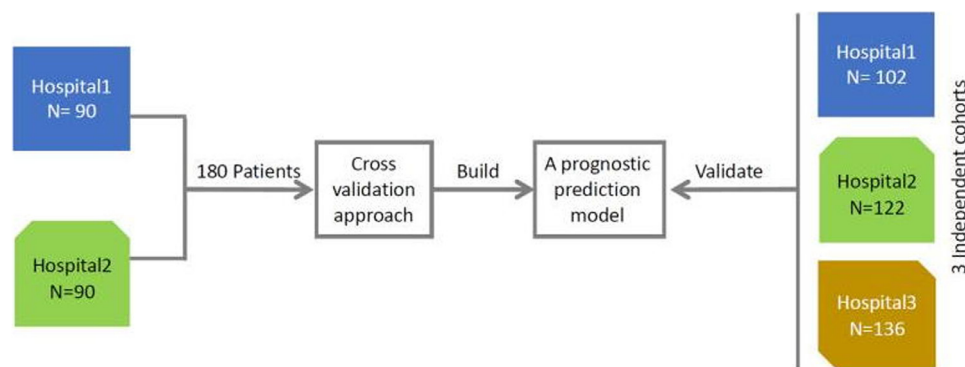

Fig. 3. A schematic diagram for study design and patient cohorts. For training set and model development, 180 patients were randomly selected from hospitals 1 and 2. The three independent hospital cohorts were used to validate the prognostic power of the score system. Hospital 1: Sir Run Run Shaw Hospital; Hospital 2: Nanjing Drum Tower Hospital; Hospital 3: Shaoxing People's Hospital.
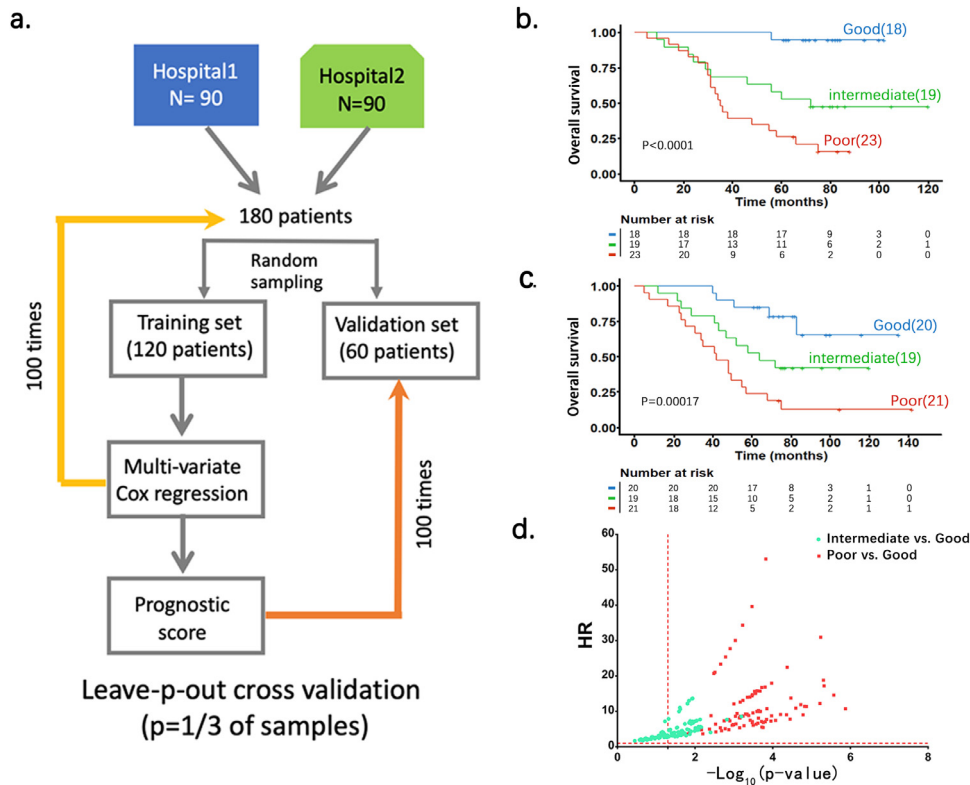
Fig. 4. The 53-gene prognostic scores predict OS of patients with gastric cancer. a. Flow diagram for building a prognostic prediction model using leave-p-out cross-validation method combined with Cox hazard regression and Kaplan-Meier analysis. b, c. Two representative Kaplan-Meier curves form the test sets on OS; d. the HR values from the test sets in Y-axis was plotted against the P values (-log$_{10}$(p-value)) from the 100 resampling validation sets in the intermediate vs. good and poor vs. good groups, respectively. Note that the red vertical line at X = 1.3 represents $P < 0.05$. The p-values were obtained by log-rank test.
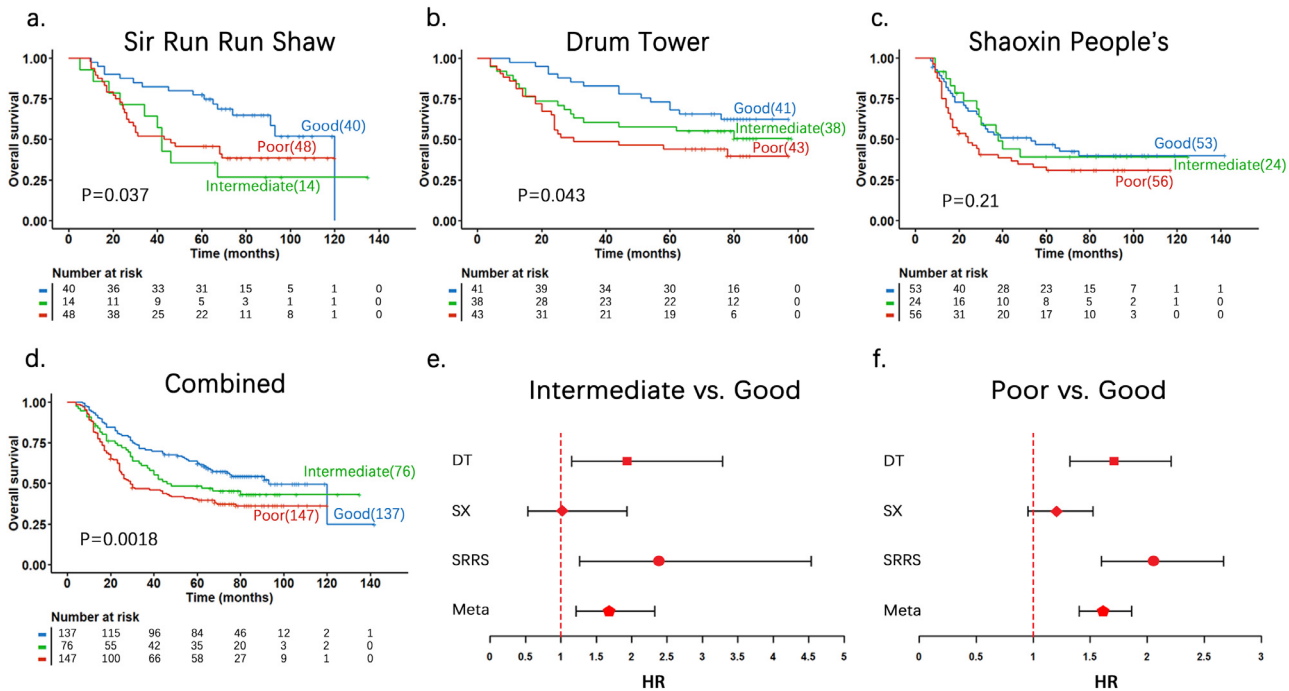

Fig. 5. Independent validation of the 53-gene prognostic score system in gastric patients from three validation cohorts and combined. a to d. Kaplan-Meier survival curves and p values for the three risk groups in the three distinct validation cohorts and combined. The p-values were obtained by log-rank test. e and f. HR and 95% CI values for different hospitals and combined. For each of 100 test sets the HR and 95% CI was calculated using a Cox model based on the prognostic score with groups. The red lines in e and f indicate a HR value of 1, or the null hypothesis. SRRS: Sir Run Run Shaw Hospital; DT: Nanjing Drum Tower Hospital, SX: Shaoxing People's Hospital.

across 100 training sets. The prognostic score was calculated as the sum of the gene expression level multiplying its averaged coefficient for each patient (see Methods, Supplementary Table 4). The patients in each resampling training set were then ranked by their scores and divided into tertile. The averaged tertile cutoff values across 100 resampling training sets were used to divide patients into good, intermediate and poor prognosis groups. K-M analysis and a log-rank test were employed to determine differences in OS between groups of patients with good, intermediate and poor prognosis in 100 resampling validation sets. Fig. 4b-c show two representative K-M survival cures from the validation sets. Fig. 4d shows the HR values from the 100 resampling validation sets for the intermediate vs. good and poor vs. good groups. We demonstrated that, compared to the patients in good prognostic group, the patients in the poor prognostic group had significantly shortened OS in all 100 validation data sets, while in the intermediate prognostic group showed no significant difference in OS in some validation data sets. These data clearly validated the 53-gene prognostic score predicting OS of gastric cancer patients.

3.2.4. Test of the 53-gene prognostic signature in independent cohorts

The prognostic prediction model developed above was then tested in three independent patient cohorts and the combined with a total of 360 patients. Prognostic score for each patient was calculated using the coefficient values obtained in the above training set. Subsequently, patients were divided into good, intermediate and poor prognostic groups based on the prognostic score using the cutoff values defined above. As shown in Fig. 5a-c, for the three individual hospital cohorts, K-M analysis demonstrated the different outcome in OS among the three groups of gastric cancer patients. Fig. 5d is the K-M curve for all three cohort patients combined ($P = 0.0018$ by log-rank test) and the distribution of 360 patients in good, intermediate or poor prognostic group was 38.1%, 21.1% and 40.8%, respectively. Fig. 5e shows the HR and 95% CI values for intermediate vs. good and Fig. 5f shows poor vs. good for the three hospital cohorts and combined. Overall, these results further demonstrated that the 53-gene prognostic scores can predict OS of gastric cancer patients.

3.2.5. Prognostic value of the 53-gene score independent of clinical variables in gastric cancer

We further analyzed whether the 53-gene prognostic score predicts prognosis in gastric cancer patients independent of clinical information including age, gender and TNM staging, WHO histologic types and differentiation (Table 2). Multivariate Cox regression analysis showed that the 53-gene prognostic score is an independent prognostic factor (Table 2). We found that the distributions of patients in the good, intermediate and poor groups are significantly different among stages (Figure 6a). The three K-M plots showed the association of the 53-gene prognostic score with OS in stage I, II, and III/IV patients (Figure 6b-d), with being significant in patients of stages II to III/IV ($P = 0.03$ and $P = 0.00057$, respectively, by log-rank test), and a trend in stage I patients ($P = 0.1$ by log-rank test). For the T grade, the percent distribution of the three prognostic groups in T1/T2, T3 and T4 was also significantly different ($P < 0.001$ by Chi-Square test) (Figure 6e), with the prognostic impact of the 53-gene signature being most significant in patients with T4 grade (Fig. 5f-h).

3.2.6. The impact of the 53-gene prognostic score on chemotherapy benefit in gastric cancer

To explore whether the 53-gene signature score was able to predict any benefit from adjuvant chemotherapy in gastric cancer patients, we stratified patients into underwent post-operative FOLFOX chemotherapy ($N = 129$) and other first-line drugs/regimens combined (other therapy, $N = 121$). Overall, the patients with good score had significantly longer OS in comparison with those with poor scores no matter what chemotherapy was used (Fig. 7). In a close

Table 2

Multivariate analysis of potential prognostic factors by Cox regression (three hospital cohort).

Variables	HR	95% CI for HR		P value
		Lower	Upper	
Clinical factors				
Age	1.017	1.006	1.028	0.003
Gender	1.138	0.841	1.539	0.403
Stage				0.091
Stage II vs. Stage I	1.526	0.527	4.417	0.436
Stage III vs. Stage I	2.660	0.732	9.664	0.137
Stage IV vs. Stage I	1.438	0.330	6.275	0.629
T grade				<0.0001
T1 vs. T4	0.584	0.199	1.709	0.326
T2 vs. T4	0.541	0.286	1.025	0.060
T3 vs. T4	0.443	0.325	0.602	<0.0001
N grade				0.208
N1 vs. N0	1.389	0.787	2.453	0.257
N2 vs. N0	1.134	0.573	2.245	0.719
N3 vs. N0	1.462	0.700	3.056	0.312
Nx vs. N0				
M grade				<0.0001
M1 vs. M0	5.416	2.479	11.835	<0.0001
Mx vs. M0	4.220	0.542	32.834	0.169
WHO classification				0.907
Well-moderate differentiation vs. well differentiation	0.567	0.238	1.351	0.200
Moderate differentiation vs. well differentiation	0.956	0.547	1.670	0.874
Poor-moderate differentiation vs. well differentiation	0.956	0.569	1.606	0.865
Poor differentiation vs. well differentiation	0.899	0.543	1.488	0.679
Undifferentiation vs. well differentiation	1.352	0.556	3.284	0.506
Adenosquamous carcinoma vs. well differentiation	2.080	0.270	16.007	0.482
Mucinous adenocarcinoma vs. well differentiation	0.943	0.493	1.803	0.859
Poorly cohesive carcinoma vs. well differentiation	1.121	0.574	2.190	0.737
Lymph node ratio	3.436	1.742	6.775	0.0003
53-Gene signature				<0.0001
Intermediate vs. Good	1.295	0.910	1.843	0.151
Poor vs. Good	2.101	1.534	2.877	<0.0001

comparison of OS between the FOLFOX and other therapy groups, we found that patients with a good score had a significantly better 5-year OS rate from FOLFOX than from other treatments. In this good score group, the 5-year OS rate reached 82% for patients underwent FOLFOX, which is significantly higher than the 5-year OS rate (61%) in the patients underwent other chemotherapies ($P = 0.028$, by Chi-Square test). We did not observe the difference in intermediate and poor score groups between the two treatment groups (Fig. 7). These results suggest that patients with a good score experienced the greatest benefit from adjuvant FOLFOX chemotherapy as compared with other chemotherapies combined.

4. Discussion

We have previously reported the identification of a 53-gene prognostic signature and scoring system that predict the OS of gastric cancer patients using publicly available multi-omics data [8]. In this study, using the TCGA-STAD dataset, we compared the prognostic ability of our signature with the 4-, 6-, and 23-gene signatures [22-24] that were published after our work, and the results clearly demonstrated that it is significantly superior to all three signatures. We also ruled out any significant correlation of our signature with certain molecular subtype(s) (i.e., EBV, MSI, GS and CIN) of gastric cancer. The latter are associated with different survival outcomes and treatment benefits [27]. To move the study of this signature to clinical validation
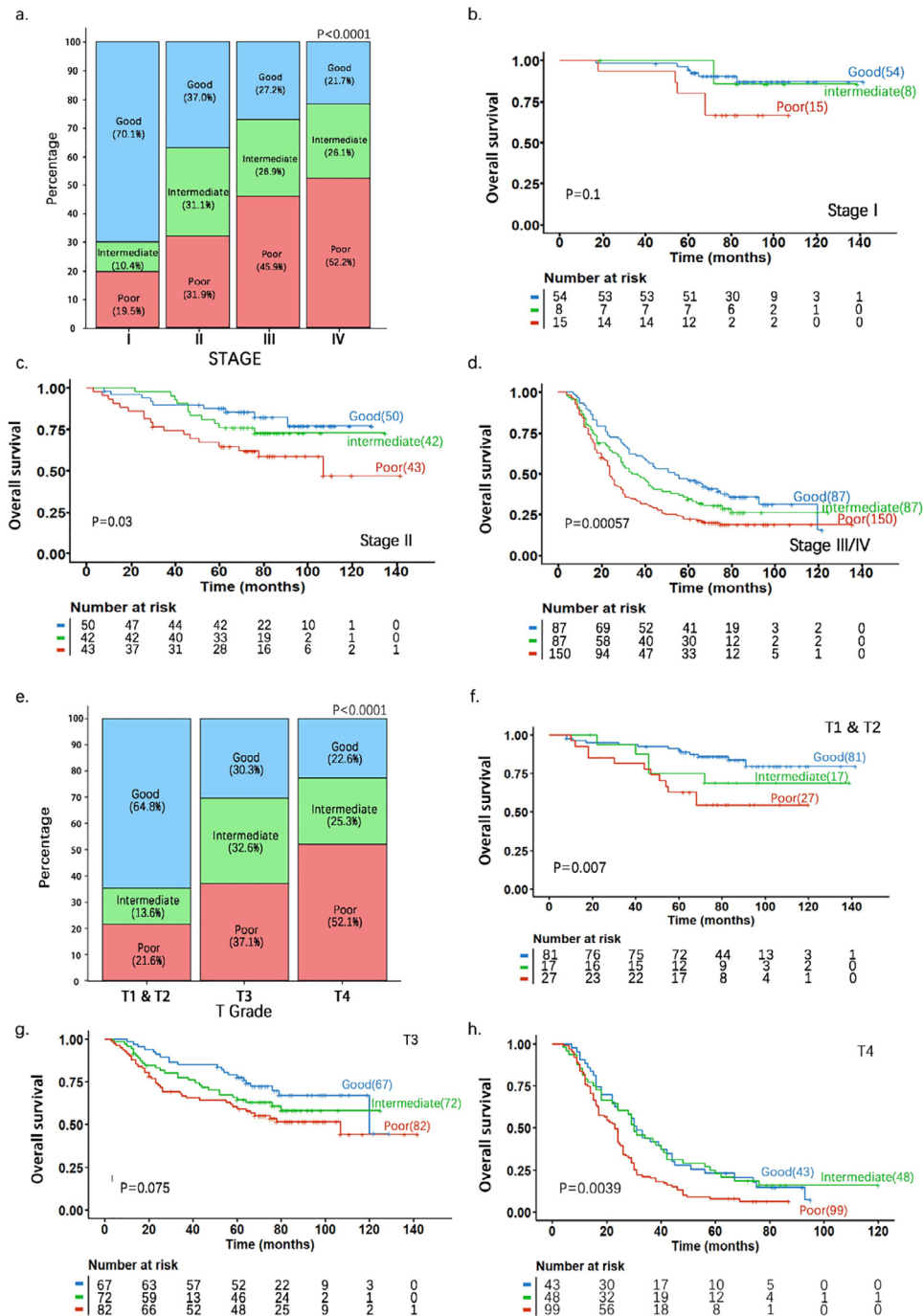
Fig. 6. The Kaplan-Meier analyses on the 53-gene score and OS in gastric cancer at different TNM stages and grades. a. Percentage distribution of three survival risk groups at different stages in pooled patient cohort. The p-value was obtained by Chi-Square test. b-d. Kaplan-Meier curves for the three risk groups on overall survival of total patients at stages I, II, and III/IV, respectively. The p-values were obtained by log-rank test. e. Percentage distribution of three survival risk groups at different T grades in pooled patient cohort. The p-value was obtained by Chi-Square test. f-h. Kaplan-Meier curves for the three risk groups on overall survival of total patients at T1&T2, T3, and T4 grades, respectively. The p-values were obtained by log-rank test.

and application, we developed an RNA hybridization-based assay for measuring gene expression in FFPE samples and then successfully confirmed the prognostic role of the score system in gastric cancer patients and also its association with disease advancement as stratified by staging in three hospitals. Given the high rates of incidence and mortality of gastric cancer in China, it is highly significance of confirming that such a score system is clinically capable of predicting OS in Chinese patients.

It is important to have a reliable, sensitive and high throughput assay for the clinical use of any prognostic gene expression detection system. Here, we developed and optimized a 53-gene expression assay for quantitative detection of mRNA in FFPE tissues of gastric cancer. FFPE is the most common format for archiving solid tissue specimens, especially tumor samples. Contrasted to a primer-probe amplicon in RT-PCR, our assay uses multiple specific short (20−30 nt) oligonucleotide probes to capture and detect target mRNAs. Hence, this assay is more suitable than RT-PCR for detecting highly degraded RNA from FFPE tissues. It is known that under the same experimental conditions, formalin fixation and RNA degradation reduce RT-PCR efficiency much more than the hybridization-based method [28].
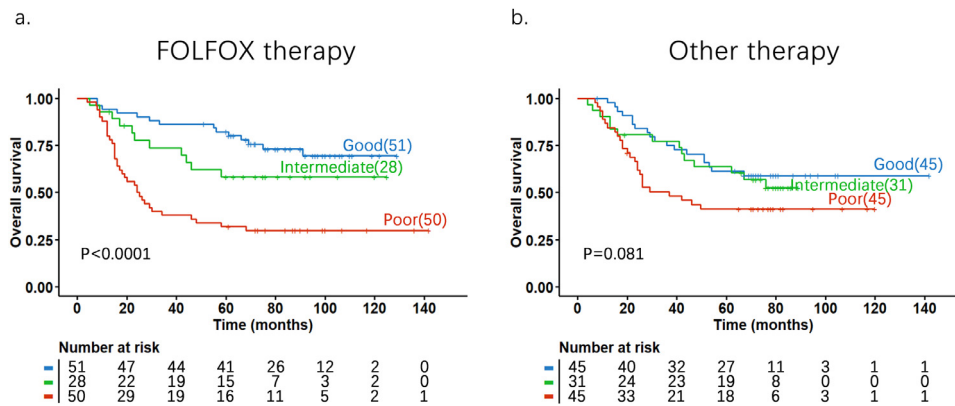
Fig. 7. Kaplan-Meier curves and P values of overall survival in two different chemotherapy groups based on 53-gene prognostic scores in gastric cancer. a. patients were treated with FOLFOX; b. patients were treated with other first-line drugs/regimes. The p-values were obtained by log-rank test.

Also, there is no enzyme(s) required in the assay as the detection is purely based on nucleic acid hybridization, which ensures no bias exists from protein activity fluctuation due to the experimental conditions. Finally, this assay is based on signal amplification rather than target amplification, which may avoid false positive results. Some previous studies have shown that this assay is more reliable, reproducible, and sensitive than qPCR in FFPE tissues [23, 25], and has also been validated using results from microarray studies [26]. Previous studies have indicated that gene expression-based biomarkers on FFPE tissues of various types of cancer are valuable for molecular classification and prognostic prediction [21, 29-31]. Here, we demonstrated that the 53-gene assay is able to measure relevant gene expression changes with high sensitivity and specificity in high-throughput platform. Therefore, this methodology is anticipated to be well suited for the clinical analysis of FFPE tissues, either in the current validation studies as well as in the future routine clinical application with the 53-gene expression detection. These results also show that our score system works in different technical platforms.

With this assay, the levels of gene expression were detected in FFPE specimens from the three hospitals, which were used to calculate a prognostic score based on the 53-gene signature and to determine a risk (low, intermediate, or high) for each patient. Through such a process, we successfully predicted in distinguishing individual patients with good prognosis from those with bad prognosis in the three hospital cohorts (102, 122, and 136 patients, respectively) and also in the combined (360 patients). As can be seen in Fig. 5d, as a whole, the long-term OS rate in the good (low-risk) group was significantly higher than that in the poor (high-risk) group ($P < 0.05$). As patients with gastric cancer are clinically categorized into four stages I to IV, we analyzed how the 53-gene score system works in each of the stage with patients from the validation set combined. We observed that, in all stages I to IV, there were clinically significant number of patients with good, intermediate and bad prognosis (Fig. 6b-d). This indicates that in every stages of gastric cancer, our score system identifies certain proportions of patients who are at high or low risk of clinical outcome. The score system identifies a gradually increased number of patients in the poor prognostic group when the staging advances, with decreasing percentage of patients in good prognosis. It can be noticed that, for stage I (Fig. 6b), patients normally bear a favorable prognosis than those with more advanced stages, the 53-gene prognostic scores still tend to predict the OS. In addition, in a multivariate Cox model, the 53-gene score provided significant prognostic power independent of patient age, gender, TNM staging and WHO classification and differentiation.

An interesting finding in this study is that our score system was able to predict adjuvant chemotherapy benefit in gastric cancer patients. Given that FOLFOX was the most popular first-line treatment regimen among all the patients enrolled in this work, we investigated whether our 53-gene score system may predict the benefit. The FOLFOX and its modified regimens have been shown to have substantial antitumor activity and tolerable toxicity for patients with advanced gastric cancer refractory to standard chemotherapy [32-34]. Although still preliminary, our data suggest that the 53-gene signature could be a promising predictive biomarker for FOLFOX regimen, at least utilized in good score (low risk)-bearing patients with different stages of gastric cancer. The clinical decision is the final judgment made by combining various genomic and clinical-histopathological factors, rather than just the routine use of a traditional NCCN-recommended chemotherapy.

This study had some limitations. First, the data analysis of specimens from three hospital cohorts was performed with totally 540 patients with gastric cancer. Although this seems to be a good sample pool, it would still lead to relatively small sample sizes when the patients were further stratified with stages and grades, as well as with different adjuvant chemotherapy regimens. Second, in this study, we used TCGA dataset and three hospital data. Due to different platform to measure the expression of 53 genes, i.e., the former was used RNA-sequencing and the latter was based on the RNA hybridization assay we developed, we had to build prognostic score separately. However, we obtained similar results and same conclusion. Third, our 53-gene assay system is based on prognostic prediction model that requires further validation in prospective studies or clinical trials, which we plan to carry out in the near future. Such studies will also be needed to further confirm the predictive potential and applicability of our 53-gene signature in the individual treatment planning with FOLFOX therapy.

In conclusion, this work primarily aimed to develop a multigene signature/score assay for prognosis of OS of gastric cancer patients in clinical use. our analysis based on a TCGA-STAD cohort showed that the 53-gene signature supersedes three previously published prognostic signatures in predicting OS of gastric cancer patients and is independent of molecular subtypes and clinical variables that are associated with patient outcomes. Using patient FFPE tumor specimens from three Chinese hospitals, our results provided evidence for the first time for the potential clinical application of the 53-gene prognostic signature in Chinese patients. The nucleic acid hybridization-based gene expression assay developed is now applicable clinically to assess the OS for gastric cancer patients. We also observed that the predictive potential of 53-gene signature-based score towards the benefit of FULFOX chemotherapy. Future prospective cohort studies with large patient sizes using this assay are warranted to fully deploy this multigene signature into clinical use.

Contributors

Study concept and design: Bo Hang, Jian-Hua Mao, Linghua Zhu, Pin Wang**;** data acquisition and analysis: Linghua Zhu, Haifeng Wang, Pin Wang, Jian-Hua Mao, Bo Hang, Chengfei Jiang, Wenhuan Li, Shuting Zhai, Dexi Jin, Xiujun Cai, Xianfa Wang, Feng Tao, Guofu Chen**;** statistical data analysis: Jian-Hua Mao, Pin Wang, Chengfei Jiang**;** manuscript drafting: Bo Hang, Jian-Hua Mao, Pin Wang**;** manuscript editing: all authors**;** funding and resource obtaining: Bin Li, Linghua Zhu, Pin Wang, Yankai Xia, Bo Hang**;** assay performance and technical support: Linghong Liao**.**

Declaration of Competing Interest

All authors declare no conflict of interest.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ebiom.2020.103023.

References

[1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394–424.

[2] Global Cancer Observatory, International agency for research on cancer. Cancer Fact Sheets, Digestive organs. Available on:https://gco.iarc.fr/today/fact-sheets-cancers.

[3] Wang FH, Shen L, Li J, et al. The chinese society of clinical oncology (CSCO): clinical guidelines for the diagnosis and treatment of gastric cancer. Cancer Commun (Lond) 2019;39:10.

[4] National Cancer Institute, Cancer Stat Facts: Stomach cancer. https://seer.cancer.gov/statfacts/html/stomach.html.

[5] Cainap C, Vlad C, Seicean A, et al. Gastric cancer: adjuvant chemotherapy versus chemoradiation. A clinical point of view. J BUON 2019;24:2209–19.

[6] Wadhwa R, Song S, Lee J, et al. Gastric cancer - molecular and clinical dimensions. Nat Rev Clin Oncol 2013;10:643–55.

[7] Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. J Clin Oncol 2005;23:7332–41.

[8] Wang P, Wang Y, Hang B, et al. A novel gene expression-based prognostic scoring system to predict survival in gastric cancer. Oncotarget 2016;7:55343–51.

[9] The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014;513:202–9.

[10] Park S, Shak S, Tang G. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817–26.

[11] van de Vijver MJ, He YD, Van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999–2009.

[12] Kratz JR, He J, Van Den Eeden SK, et al. A practical molecular assay to predict survival in resected non-squamous, non-small cell lung cancer: development and international validation studies. Lancet 2012;379:823–32.

[13] Clark-Langone KM, Sangli C, Krishnakumar J, et al. Translating tumor biology into personalized treatment planning: analytical performance characteristics of the oncotype DX colon cancer assay. BMC Cancer 2010;10:691.

[14] Chen EG, Wang P, Lou H, et al. A robust gene expression-based prognostic risk score predicts overall survival of lung adenocarcinoma patients. Oncotarget 2017; 9:6862–71.

[15] Zhao Y., Yang S.M., Jin Y.L., et al. A robust gene expression prognostic signature for overall survival in high-grade serous ovarian cancer. J Oncol 2019:3614207.

[16] Huet S, Tesson B, Jais JP, et al. A gene-expression profiling score for prediction of outcome in patients with follicular lymphoma: a retrospective training and validation analysis in three international cohorts. Lancet Oncol 2018;19:549–61.

[17] Ooi CH, Ivanova T, Wu J, et al. Oncogenic pathway combinations predict clinical prognosis in gastric cancer. PLoS Genet 2019;5:e1000676.

[18] Kim H, Eun JW, Lee H, et al. Gene expression changes in patient-matched gastric normal mucosa, adenomas, and carcinomas. Exp Mol Pathol 2011;90:201–9.

[19] Nam S, Lee J, Goh SH. Differential gene expression pattern in early gastric cancer by an integrative systematic approach. Int J Oncol 2012;41:1675–82.

[20] Yin Y, Zhuo W, Zhao Y, et al. Converting a microarray signature into a diagnostic test: a trial of custom 74 gene array for clarification and prediction the prognosis of gastric cancer. PLoS ONE 2013;8:e81561.

[21] Zhao X, Cai H, Wang X, et al. Discovery of signature genes in gastric cancer associated with prognosis. Neoplasma 2016;63:239–45.

[22] Lee KW, Lee SS, Hwang JE, et al. Development and validation of a six-gene recurrence risk score assay for gastric cancer. Clin Cancer Res 2016;22:6228–35.

[23] Wang X, Liu Y, Niu Z, et al. Prognostic value of a 25-gene assay in patients with gastric cancer after curative resection. Sci Rep 2017;7:7515.

[24] Cheong JH, Yang HK, Kim H, et al. Predictive test for chemotherapy response in resectable gastric cancer: a multi-cohort, retrospective analysis. Lancet Oncol 2018;19:629–38.

[25] Knudsen BS, Allen AN, McLerran DF, et al. Evaluation of the branched-chain DNA assay for measurement of RNA in formalin-fixed tissues. J Mol Diagn 2008;10:169–76.

[26] Canales RD, Luo Y, Willey JC, et al. Evaluation of DNA microarray results with quantitative gene expression platforms. Nat Biotechnol 2006;24:1115–22.

[27] Sohn BH, Hwang JE, Jang HJ, et al. Clinical significance of four molecular subtypes of gastric cancer identified by the cancer genome atlas project. Clin Cancer Res 2017;23:4441–9.

[28] Yang W, Maqsodi B, Ma Y, et al. Direct quantification of gene expression in homogenates of formalin-fixed, paraffin-embedded tissues. BioTechniques 2006;40:481–6.

[29] Ferrer M Cormeo B, Davis J, et al. A multiplex high-throughput gene expression assay to simultaneously detect disease and functional markers in induced pluripotent stem cell-derived retinal pigment epithelium. Stem Cells Transl Med 2014;3:911–22.

[30] Rini B, Goddard A, Knezevic D, et al. A 16-gene assay to predict recurrence after surgery in localised renal cell carcinoma: development and validation studies. Lancet Oncol 2015;16:676–85.

[31] Scott DW, Chan FC, Hong F, et al. Gene expression-based model using formalin-fixed paraffin-embedded biopsies predicts overall survival in advanced-stage classical Hodgkin lymphoma. J Clin Oncol 2013;31:692–700.

[32] Yeh YS, Tsai HL, Ma CJ, et al. A retrospective study of the safety and efficacy of a first-line treatment with modified FOLFOX-4 in unresectable advanced or recurrent gastric cancer patients. Chemotherapy 2012;58(5):411–8.

[33] Catalano V, Bisonni R, Graziano F, et al. A phase II study of modified FOLFOX as first-line chemotherapy for metastatic gastric cancer in elderly patients with associated diseases. Gastric Cancer 2013;16(3):411–9.

[34] Mitani S, Kadowaki S, Komori A, et al. A phase ii study of modified FOLFOX6 for advanced gastric cancer refractory to standard therapies. Adv Ther 2020 May 7.