A long non-coding RNA signature to improve prognosis prediction of gastric cancer

Xiaoqiang Zhu†, Xianglong Tian†, Chenyang Yu†, Chaoqin Shen†, Tingting Yan†, Jie Hong*, Zheng Wang*, Jing-Yuan Fang* and Haoyan Chen*

Abstract

Background: Increasing evidence suggests long non-coding RNAs (lncRNAs) are frequently aberrantly expressed in cancers, however, few related lncRNA signatures have been established for prediction of cancer prognosis. We aimed at developing lncRNA signature to improve prognosis prediction of gastric cancer (GC).

Methods: Using a lncRNA-mining approach, we performed lncRNA expression profiling in large GC cohorts from Gene Expression Ominus (GEO), including GSE62254 data set (N = 300) and GSE15459 data set (N = 192). We established a set of 24-lncRNAs that were significantly associated with the disease free survival (DFS) in the test series.

Results: Based on this 24-lncRNA signature, the test series patients could be classified into high-risk or low-risk subgroup with significantly different DFS (HR = 1.19, 95% CI = 1.13–1.25, P < 0.0001). The prognostic value of this 24-lncRNA signature was confirmed in the internal validation series and another external validation series, respectively. Further analysis revealed that the prognostic value of this signature was independent of lymph node ratio (LNR) and postoperative chemotherapy. Gene set enrichment analysis (GSEA) indicated that high risk score group was associated with several cancer recurrence and metastasis associated pathways.

Conclusions: The identification of the prognostic lncRNAs indicates the potential roles of lncRNAs in GC biogenesis. Our results may provide an efficient classification tool for clinical prognosis evaluation of GC.

Keywords: Gastric cancer, LncRNAs, LNR, GSEA, Survival

Background

Being the fourth most common malignancy, GC has been the second leading cause of cancer deaths worldwide [1]. An estimated 951,600 GC cases occurred and 723,100 patients died from GC in 2012 [1, 2]. Adequate surgical resection is the only curative therapeutic option for GC [3, 4]. The current strategy to GC management, which has significantly improved overall survival (OS) [4], includes endoscopic detection followed by gastrectomy and chemotherapy or chemo-radiotherapy in neoadjuvant or adjuvant regiments [5]. However, treatment outcome still remains undesirable. The current Union International Committee on Cancer (UIICC) or the American Joint Committee on Cancer (AJCC) TNM stage system has shown valuable but insufficient prediction for prognosis and estimation for subsets of GC patients [6–8]. An increasing amount of evidence demonstrates that the discovery and application of molecular biomarkers will promote the prognostic evaluation and identification of potential high-risky GC patients [5, 9, 10].

Currently, with the advancements in transcriptome profiling, the roles of dysregulated functional long non-coding RNAs (lncRNAs) in human cancers have received considerable attention [11–13]. LncRNAs are mRNA-like transcripts ranging in length from 200 nucleotides (nt) to ~ 100 kilobases (kb) that lack significant
protein-coding abilities [14, 15]. Increasing evidence suggests that the aberrant expressions of lncRNAs have been associated with human cancers [16–18], and some of them have been implicated in diagnosis and prognosis [19, 20]. Although several prognostic biomarkers for GC have been undergoing or tested in clinical trials such as Fibroblast Growth Factor Receptor (FGFR) [21], Human Epidermal Growth Factor Receptor 2 (HER2) [22], Epidermal Growth Factor Receptor (EGFR) [23], Hepatocyte Growth Factor Receptor (HGFR) [24], etc, many more potential and valuable molecular biomarkers are urgent to be discovered and identified to improve the clinical outcome of patients with GC. Increasing studies have shown that lncRNAs could be one of the best candidates as potential prognostic biomarkers in GC [25–27]. Therefore, searching a lncRNA signature might be concrete predictive and prognostic value in the management of GC.

However, lncRNA profiles in most human cancers remain largely unknown, mainly due to the lack of such arrays. Previous studies have demonstrated that lncRNA profiling could be achieved by mining previously published gene expression microarray data because a large amount of lncRNA-specific probes were fortuitously represented on the commonly used microarray platforms [28, 29]. In the present study, we applied this method to conduct gene expressions of lncRNAs profiling on a cohort of 300 patients from GSE62254 as well as another independent data set from GEO database. By using the sample-splitting method, random survival forests-variable hunting (RSF-VH) algorithm and Cox regression analysis, we identified a prognostic, 24-lncRNA signature from the GSE62254 test series patients, and validated it in the GSE62254 validation series and another independent GEO cohort (GSE15459).

**Methods**

**GC datasets preparation**

Microarray data from GSE62254 and GSE15459 data sets were directly downloaded from Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/). These data sets corresponded to all available public datasets fulfilling the following criteria: available gene expression data obtained using the same chip platform (Affymetrix Human Genome U133 Plus 2.0 chips) with raw data CEL files, and patient outcome data available. After initial quality check, two panels of GC gene expression data sets were included in our study: GSE62254 and GSE15459. The GC samples in GSE62254 and GSE15459 were randomly split into a test series (n = 180) and an internal validation series (n = 120). Additionally, the GC samples in GSE15459 were analyzed as an external validation series.

**Microarray data processing and lncRNA profile mining**

The raw CEL files were downloaded from GEO database and background adjusted using Robust Multichip Average (RMA) [30] which has been shown to be a solid measure tool for lncRNA profiling data [31]. The approach of lncRNA profile mining mainly referred to Xiaoqin Zhang et al [32]. Briefly, we mapped the Affymetrix HG-U133 Plus 2.0 probe set IDs to the NetAffx Annotation Files. Based on the Refseq transcript ID and/or Ensembl gene ID in NetAffx annotations, we only retained non-coding protein genes and further filtered them by eliminating pseudogenes including microRNAs, rRNAs and other short RNAs such as snoRNAs, snRNAs and tRNAs. Finally, 2448 annotated lncRNA transcripts with corresponding Affymetrix probe IDs were generated.

**GSEA**

GSEA was performed by the JAVA program (http://software.broadinstitute.org/gsea/index.jsp) using MSigDB C2 CP: Canonical pathways gene set collection. The GSEA, visualized in Cytoscape (version 2.8.0) was used to determine whether the members of a given gene set were generally associated with risk score, and was therefore conducted on all mRNA genes on the HG-U133 Plus 2.0 ranked by enrichment score from most positive and most negative. 1000 random sample permutations were carried out, and the significance threshold set at FDR < 0.01. If a gene set had a positive enrichment score, the majority of its members had higher expression accompanied with higher risk score, and the set was termed “enriched”.

**Bioinformatics analysis**

All statistical analyses were conducted using R software [33] and Bioconductor [34]. The association between the lncRNA expression and patient’s DFS or OS was assessed by univariable Cox regression analysis along with a permutation test using BRB-ArrayTools [35]. The permutation p-values for significant genes were computed based on 10,000 random permutations and genes were considered statistically significant if their permutation p values were less than 0.01. And genes that passed the filter criteria were considered for further analysis by applying the random survival forest-variable hunting (RSF-VH) algorithm [36]. Among the parameters involving in this algorithm, the number of nsplit was set as nsplit = 10 following Ishwaran and colleagues [37] in the variable selection function within the Random Survival Forest package during the selection. To construct a predictive model, the candidate genes were fitted in a univariable Cox regression model in the test series as previously applied [38]. A risk score formula was then established by including each of these selected genes, weighted by their estimated regression coefficients in the univariable
Cox regression analysis [38]. With this risk score formula, patients in each set were classified into high-risk or low-risk group by using the corresponding median risk score as the cutoff point. Survival differences between the high-risk and low-risk groups in each set were assessed by the Kaplan-Meier estimate, and compared using the log-rank test. To test whether the risk score was independent of LNR and postoperative chemotherapy, multivariable Cox regression and data stratification analysis were performed. We performed ROC analysis to compare the sensitivity and specificity of the survival prediction based on the IncRNA risk score, AJCC stage, LNR and postoperative chemotherapy. To generate ROC curves, patients were classified as surviving either longer or shorter than the median DFS, excluding patients who were alive for durations less than the median DFS at last follow-up [39]. In the log-rank test, Cox regression analysis and ROC analysis, the significance was defined as $P$ values being less than 0.05.

**Results**

**GC data sets preparation**

GC data sets and corresponding clinical data were downloaded from the publicly available GEO database. The following two cohorts of GC gene expression data were included in this study: GSE62254 [5] and GSE15459 [40]. After removal of the samples without survival status, a total of 492 GC patients analyzed in the present study (see Additional file 1). These included 300 GC patients from GSE62254 (180 patients from the test series and 120 patients from the validation series). And 192 GC patients from GSE15459 were included after 8 patients were removed due to absence of clinical outcome information.

**Identification of prognostic IncRNA genes from the test series**

The 300 GC samples were randomly assigned to a test series ($n = 180$) or a validation series (120). The test series was used for the detection of prognostic IncRNA genes. By subjecting the IncRNA expression data of the test series to univariable Cox regression proportional hazards regression analysis using Biometric Research Branch-Array (BRB-Array) Tools, we identified a set of 63 genes whose parameter $P$-value were less than 0.01. Those 63 genes were further analyzed by random survival forest-variable hunting (RSF-VH) algorithm [36]. This algorithm is a high-dimensional order statistic measuring the predictiveness of a variable in a survival tree that exploits maximal subtrees for effective variable selection under such scenarios [36]. With this method, 24 genes were screened out as the predictors (genes). Table 1 showed a list of genes with their obtained specific values including permutation $P$ values, hazard ratios and coefficients which of these were derived from the univariable Cox proportional hazards regression analysis. Moreover, the variable importance values were also figured out following the variable selection function within the Random Survival Forest package. Variable importance measures the increase (or decrease) in the prediction error for the random forests model when a variable is randomly “noise up”. That is if the prediction error of the model became worse when the effect of a variable in the model on the prediction was intentionally destroyed, this means that the variable is important in the model [41, 42]. Among these genes, positive coefficients indicated that the higher expression levels of 14 genes (AF035291, AI028608, AK026189, H04858, BC037827, BC038210, AI916498, AA463827, AA041523, BE621082, AK056852, AW26234, AL703532, AI905542) were associated with shorter survival. The negative coefficients for the remaining ten genes (AI080288, BC021187, BF238392, BC005107, BC039674, AI056187, T79746, H11436, BF511694, BC035722) indicated that their higher levels of expression were associated with longer survival.

**The 24-IncRNA signature and the patients’ survival in the test series**

A risk-score formula was created based on the expression of these 24 IncRNAs for DFS prediction, as follows: Risk score = $(2.11846\times$expression level of AF035291) + (1.92247$\times$expression level of AI028608) + (1.53266$\times$expression level of AK026189) + (1.3926$\times$expression level of H04858) + (1.27718$\times$expression level of BC037827) + (1.20171$\times$expression level of BC038210) + (1.04591$\times$expression level of AI916498) + (1.0294$\times$expression level of AA463827) + (0.92436$\times$expression level of AA041523) + (0.81047$\times$expression level of BE621082) + (0.65369$\times$expression level of AK056852) + (0.54056$\times$expression level of AW26234) + (0.2811$\times$expression level of AL703532) + (0.24825$\times$expression level of AI905542) + (-1.86125$\times$expression level of AI080288) + (-2.24862$\times$expression level of BC021187) + (-2.61423$\times$expression level of BF238392) + (-2.65478$\times$expression level of BC035722) + (-2.97863$\times$expression level of BC039674) + (-3.31933$\times$expression level of BF511694) + (-3.40924$\times$expression level of BC035722). We then calculated the 24-IncRNA signature risk score for each patient in the test series, and ranked them according to their risk scores. As such, patients were divided into a high-risk group ($n = 90$) or a low-risk group ($n = 90$) using the median risk score of the test series as the cutoff point. Patients in the high-risk group had significantly shorter median DFS than those in the low-risk group (log-rank test $P < 0.0001$) (Fig. 1a). The association of the 24-IncRNA risk score with DFS was also significant when it was evaluated as a continuous variable in the multivariable Cox regression model (Fig. 2a).
Validation of the 24-lncRNA signature for survival prediction in the validation series and the entire GSE62254 data set

To confirm our findings, we validated our 24-lncRNA signature in the internal validation series. By using the same risk formula, we classified patients into high-risk \((n=60)\) or low-risk group \((n=60)\) using the median score of the internal validation series as the cutoff point. In accordance with the findings described above, patients in the high-risk group had significantly shorter median DFS than those in the low-risk group (log-rank test \(P=0.0126\)) (Fig. 1b). Risk score-based classification of the entire GSE62254 cohort (i.e. combined test and validation series) also yielded similar results (log-rank test \(P<0.0001\)) (Fig. 1c). In the multivariable Cox regression model that the 24-lncRNA risk score was evaluated as a continuous variable, similar correlation could be observed (Fig. 2b-c).

Table 1 LncRNAs significantly associated with the disease-free survival in the test series patients \((N=180)\)

| Probe     | Gene symbol | Permutation P valuea,b | Hazard ratioa | Coefficienta | VI | Associated diseases                                                                 | Description                                                                                           |
|-----------|-------------|------------------------|---------------|--------------|----|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|
| 233512_at | AF035291    | 0.0017                 | 8.325         | 2.11846      | 0.012 | NR                                                                                   | hypothetical LOC100287216                                                                               |
| 1568854_at| AI028608    | 0.0084                 | 6.839         | 1.92247      | −0.002 | NR                                                                                   | non-protein coding RNA 240                                                                            |
| 229280_s_at| AK026189    | 6.00E-04               | 4.633         | 1.53266      | 0.006 | Melanoma and basal cell carcinoma                                                   | promotes metastasis and recurrence of melanoma                                                          |
| 1559412_at| H04858      | 8.00E-04               | 4.030         | 1.3926       | 0.002 | Acute megakaryoblastic leukemia                                                      | Maintenance of leukemic growth                                                                         |
| 1559965_at| BC037827    | 0.0018                 | 3.587         | 1.27718      | 0.053 | NR                                                                                   | hypothetical LOC100192378                                                                               |
| 1557338_s_at| BC038210   | 0.0108                 | 3.326         | 1.20171      | 0.017 | NR                                                                                   | NR                                                                                                    |
| 230589_s_at| AI916498    | 0.0034                 | 2.844         | 1.04591      | −0.006 | Gastric cancer                                                                       | Sensitivity of iodine-125 particle irradiation and regulation of NF-KB signaling pathway              |
| 239466_at | AA463827    | 0.0013                 | 2.799         | 1.0294       | 0.026 | NR                                                                                   | hypothetical LOC344595                                                                                |
| 230251_at | AA041523    | 0.0014                 | 2.521         | 0.92463      | 0.038 | Non-small cell lung cancer                                                          | Associated with loss function of LKB1 gene                                                             |
| 225029_at | BE621082    | 0.0016                 | 2.249         | 0.81047      | 0.007 | Systemic lupus erythematosus                                                        | A novel susceptibility locus on Xp11.21                                                               |
| 1564139_at| AK056852    | 0.0016                 | 1.923         | 0.65369      | 0.004 | NR                                                                                   | hypothetical LOC144571                                                                                |
| 229014_at | AW206234    | 0.0021                 | 1.718         | 0.54056      | 0.002 | NR                                                                                   | hypothetical LOC441094                                                                                |
| 1558828_s_at| AL703532   | 0.0032                 | 1.325         | 0.2811       | 0.005 | Heart failure associated diseases                                                   | Regulation of cardiac cell differentiation and homeostasis                                             |
| 235759_at | AI095542    | 0.0034                 | 1.282         | 0.24825      | −0.004 | NR                                                                                   | NR                                                                                                    |
| 213972_at | AI080288    | 0.0036                 | 0.155         | −1.86125     | −0.007 | NR                                                                                   | NR                                                                                                    |
| 1554880_at| BC021187    | 0.0024                 | 0.106         | −2.24862     | −0.002 | Gastric cancer                                                                       | Protective roles for gastric cancer                                                                   |
| 235824_at | BF238392    | 0.0019                 | 0.073         | −2.61423     | 0.012 | NR                                                                                   | NR                                                                                                    |
| 232191_at | BC005107    | 0.0076                 | 0.070         | −2.65478     | 0.002 | chromosome 21 open reading frame 105                                               | NR                                                                                                    |
| 1563296_at| BC039674    | 0.0022                 | 0.068         | −2.69258     | 0.000 | NR                                                                                   | NR                                                                                                    |
| 239617_at | AI056187    | 9.00E-04               | 0.061         | −2.79863     | −0.001 | Gastric cancer                                                                       | spans the promoter and untranslated regions of the ghrelin gene                                       |
| 243975_at | T79746      | 0.0031                 | 0.058         | −2.85076     | 0.003 | NR                                                                                   | NR                                                                                                    |
| 1558666_at| H11436      | 3.00E-04               | 0.055         | −2.89127     | −0.007 | NR                                                                                   | NR                                                                                                    |
| 237471_at | BF511694    | 0.0034                 | 0.041         | −3.19733     | 0.003 | NR                                                                                   | hypothetical LOC154822                                                                                |
| 1562683_a_at| BC035722   | 0.0023                 | 0.033         | −3.40924     | 0.001 | NR                                                                                   | hypothetical LOC285547                                                                                |

Abbreviations: VI Variable Importance, NR Not Reported

aDerived from the univariable Cox proportional hazards regression analysis in the 180 test series patients

bObtained from permutation test repeated 10,000 times
with low-risk scores tended to express high level of protective lncRNAs (AI080288, BC021187, BF238392, BC005107, BC039674, AI056187, T79746, H11436, BF511694, BC035722).

Further validation of the 24-lncRNA signature in another independent data set
We further validated our 24-lncRNA signature in another independent GC data set obtained from GEO, GSE15459. The clinical characteristics of this cohort were also listed (see Additional file 1). Although the patient outcome was represented with OS, this data set confirmed the ability of our model in predicting survival. As shown in Fig. 1d, the 24-lncRNA model could effectively predict the OS in patients from GSE15459 (log-rank test \( P = 0.0084 \)). In the multivariable Cox regression model, the lncRNA risk score was significantly associated with OS as a continuous variable in the GSE15459 cohort (Fig. 2d).

**Prognostic value of the 24-lncRNA signature is independent of LNR**

LNR is the ratio of the numbers of metastatic lymph nodes to those of the dissected lymph nodes. Increasing evidence indicated that LNR is a novel and simple marker which can easily stratify the prognoses of advanced GC [43–45]. And several studies have demonstrated that LNR = 16.7 % was the optimal cutoff level as an effective prognostic indicator in advanced GC [46, 47]. Fortunately, LNR could be calculated out in GSE62254 data set for 300 patients. Thus, we tested whether the prognostic value of the 24-lncRNA signature was independent of LNR. For this, we first conducted multivariable Cox regression analysis and stratification analysis. In the multivariable Cox regression model, the lncRNA risk score was significantly associated with OS as a continuous variable in the GSE15459 cohort (Fig. 2d).
stratification analysis was then performed which stratified these patients into LNR $\geq 16.7\%$ subgroup and LNR $< 16.7\%$ subgroup. The stratification analysis showed that the 24-lncRNA signature could identify patients with different prognoses despite of the same LNR stratum (Fig. 4a). For instance, among the patients LNR $\geq 16.7\%$ ($n=139$), the 24-lncRNA risk score could further subdivide them into those likely to have longer versus shorter survival (log-rank test $P < 0.0001$) (Fig. 4b). Similarly, among those LNR $< 16.7\%$ ($n=161$), the 24-lncRNA risk score could also subdivide patients into two subgroups with significantly disparate survival (log-rank test $P < 0.0001$) (Fig. 4c).

Prognostic value of the 24-lncRNA signature is independent of postoperative chemotherapy

We also tested whether the prognostic value of the 24-lncRNA signature was independent of postoperative chemotherapy. To achieve this, we first conducted multivariable Cox regression analysis and stratification analysis. By inspection, of the 300 GSE62254 samples analyzed, 299 patients had available data on their postoperative chemotherapy. Unfortunately, of the 192 patients from GSE15459, no patients had available information on their postoperative chemotherapy. In the multivariable Cox regression analysis on these 299 patients that contained 24-lncRNA risk score, postoperative chemotherapy, age and gender as
covariates, we found that the 24-lncRNA risk score (HR = 1.17, 95 % CI = 1.13–1.22, \( P < 0.0001 \)) and postoperative chemotherapy (HR = 0.38, 95 % CI = 0.19–0.76, \( P = 0.0060 \)) were both independent prognostic factors (Table 2). Data stratification analysis was then performed which stratified these patients into with postoperative chemotherapy subgroup or without postoperative chemotherapy subgroup. The stratification analysis showed that the 24-lncRNA signature could identify patients with different prognoses despite of the same postoperative chemotherapy stratum (Fig. 5a). For instance, among the patients with postoperative chemotherapy (\( n = 80 \)), the 24-lncRNA risk score could further subdivide them into those likely to have longer versus shorter survival (log-rank test \( P = 0.0007 \)) (Fig. 5b). Similarly, among those without postoperative chemotherapy (\( n = 219 \),...
Table 2: Multivariable Cox regression analysis of the 24-lncRNA risk score, LNR and postoperative chemotherapy in GSE62254 series

| Variables                     | HR   | 95 % CI of HR | P value |
|-------------------------------|------|---------------|---------|
| 24-lncRNA risk score (N = 300) | 1.16 | 1.10 – 1.21   | <0.0001 |
| LNR                           | 4.33 | 3.14 – 1.93   | 0.014   |
| Age                           | 1.03 | 1.00 – 1.05   | 0.032   |
| Gender                        | 1.00 | 0.65 – 1.84   | 0.732   |
| TNM stage I                   | 1.00 | (Referent)    |         |
| TNM stage II                  | 1.08 | 0.23 – 5.08   | 0.919   |
| TNM stage III                 | 2.09 | 0.47 – 9.24   | 0.330   |
| TNM stage IV                  | 3.89 | 0.82 – 18.44  | 0.087   |
| 24-lncRNA risk score (N = 299) | 1.14 | 1.09 – 1.19   | <0.0001 |
| Postoperative chemotherapy    | 0.36 | 0.18 – 0.73   | 0.004   |
| Age                           | 1.01 | 0.99 – 1.03   | 0.324   |
| Gender                        | 1.04 | 0.62 – 1.75   | 0.897   |
| TNM stage I                   | 1.00 | (Referent)    |         |
| TNM stage II                  | 1.47 | 0.31 – 6.85   | 0.626   |
| TNM stage III                 | 4.00 | 0.93 – 17.15  | 0.062   |
| TNM stage IV                  | 8.50 | 2.00 – 36.20  | 0.004   |

In Cox regression analysis, risk score, LNR and age were evaluated as continuous variables, and postoperative chemotherapy and gender were evaluated as category variables.
Abbreviations: LNR lymph node ratio, HR hazard ratio.

the 24-lncRNA risk score could still subdivide patients into two subgroups with significantly disparate survival (log-rank test P < 0.0001) (Fig. 5c).

Prognostic value of the 24-lncRNA signature is independent of TNM stage

According to TNM stage system for GC, patients in GSE62254 series were divided into four subgroups (I, II, III and IV). The stratification analysis suggested that the 24-lncRNA signature could identify patients with different prognoses in each TNM stage subgroup (Fig. 6a-d) despite that the P value was not significant in stage I (log-rank test P = 0.2900). This might be because the sample size was too small (only 30 patients, Fig. 6a) to draw any reliable conclusions. Interestingly, when low TNM stage (I & II) and high TNM stage (III & IV) were combined, respectively, the 24-lncRNA signature could also identify patients with different prognoses in each subgroup and the P value was significant (see Additional file 2).

Additionally, we performed ROC analysis to compare the sensitivity and specificity of survival prediction among the 24-lncRNA risk score model, AJCC stage, LNR and postoperative chemotherapy. The area under receiver operating characteristic (AUROC) was assessed and compared among the four prognostic factors. As shown in Fig. 7, there was no significant difference between the AUROC of 24-lncRNA risk score when compared with AJCC stage (0.82 versus 0.76, 95 % CI = 0.76–0.88, P = 0.1861). However, the AUROC of the 24-lncRNA signature risk score combined with AJCC stage was significantly greater than AJCC stage alone (0.85 versus 0.76, 95 % CI = 0.69–0.83, P = 0.0002). Additionally, the 24-lncRNA signature risk score was significantly superior than that of LNR (0.82 versus 0.71, 95 % CI = 0.62–0.79, P = 0.0297) and postoperative chemotherapy (0.82 versus 0.63, 95 % CI = 0.55–0.70, P < 0.0001). Although the predictive ability of the 24-lncRNA signature was equivalent to AJCC stage, these results also indicated that the 24-lncRNA signature combined with AJCC stage may have a stronger power for DFS prediction in the ROC analysis. Also, the 24-lncRNA signature may have a better predictive ability than both LNR and postoperative chemotherapy alone.

Identification of 24-lncRNA signature correlated biological pathways and processes

We performed GSEA to identify correlated biological process and signaling pathways using the 24-lncRNA signature on the basis of risk score for classification. Significant gene sets (FDR < 0.001) were visualized as interaction networks with Cytoscape (Fig. 8a, see Additional file 3). The high risk score was accompanied with up-regulation of several cancer-related networks including recurrence, metastasis and cancer stemness associated pathways. For instance, Polo-like kinase 1 (PLK1) and E2F-mediated associated pathways were implicated in cancer recurrence and metastasis [48–50]. We proposed that the 24-lncRNA signature might be involved in these networks. Since cancer recurrence and metastasis could strongly affect patients’ DFS, we compared the risk score of patients with recurrence and without recurrence (non-recurrence) in GSE62254 series when this information was available. Patients with recurrence tended to have higher risk score than patients without recurrence (Fig. 8b, P < 0.0001).

Discussion

Currently, the discovery of thousands of IncRNAs has broken the conventional thinking that the gene regulation in biology was mostly involved in protein-coding genes [51, 52]. Evidence from growing publications have demonstrated that functional IncRNAs expression patterns were associated with human cancers [11–14]. These IncRNAs were implicated in various tumorigenesis processes including proliferation [53], invasion [54] and apoptosis [55] by acting as tumor oncogenes or suppressors. The aberrant expressions of specific IncRNAs in cancer can mark the spectrum of disease progression and may serve as independent biomarkers for diagnosis and prognosis [29, 32]. More recently, IncRNAs have been associated with biology of GC. However, the prognostic values of IncRNAs in GC have not been clarified clearly.
To identify the prognostic lncRNA genes, we profiled lncRNA by mining the existing microarray gene expression data on a variety of commonly used commercial arrays. Of those, the Affymetrix Human Genome U133 array series is one of the most commonly used commercial microarrays in human cancer profiling [56]. As a public gene expression data repository, GEO has contained lots of gene expression data that could be used for further analysis. Based on this mining method, we additionally applied another method to select prognostic lncRNA genes. Predictors (genes) were selected by applying the random survival forest-variable hunting (RSF-VH) algorithm [36]. The random forests method is classified into a tree-based method which has an advantage in detecting interactions. This algorithm exploits maximal subtrees for effective variable selection, and the trees in a survival forest are grown randomly using a two-step randomization process [36]. Moreover, it has been developed for processing data with several variables larger than the number of samples. There is no denying that many published studies applied univariable and multivariable analyses on microarray data for screening where potential genes interacting with other genes may be dropped from the analyses. Actually, in this regard, the RSF-VH algorithm would be more powerful.

Functional characteristics of the 24 prognostic lncRNAs

We finally identified a set of 24 lncRNAs that showed differential expressions among the GC patients included in the data sets. Such differentiations signified their potential roles in GC. Although some of these deregulated lncRNAs have been reported to express in cancer or other disorders, they have not been investigated in GC. For example, the expression of AK026189 (CASC 15) was found to be associated with neuroblastoma and was increased during melanoma progression [57–59]. And it was regarded as an independent predictor of disease recurrence in a cohort of 141 patients with AJCC stage III lymph node metastasis [58]. In our study, AK026189 was highly expressed in GC and was found to be correlated with shortened survival. Another candidate, H048558 (MIR99AHG, MONO) was also abundantly expressed in
GC samples. A study has revealed that H04858 was highly expressed in acute megakaryoblastic leukemia cell lines serving as a regulator of hematopoiesis and oncogene in the development of myeloid leukemia [60]. Thus, we infer that H04858 may act as an oncogene in GC tumorigenesis and further investigations are great needed as well.

Moreover, lncRNA AI916498 (TRAF3IP2-AS1) was found differentially expressed in midbrain dopamine cells of human cocaine abusers and its transcript showed a surprisingly strong nuclear localization in dopamine cells [61]. Bannon et al [61] suggested that AI916498 might act as a mediator of a disruption of NF-KB signaling seen in cocaine abuse. More interestingly, AI916498 was down-regulated in the human gastric cell lines after received iodine-125 particle irradiation [62]. This indicated that AI916498 may play a critical role in the iodine-125 seed treatment of GC and be a potential target for developing anti-gastric cancer drugs in the future. Also, AA041523 (LINC00473, C6orf176) was first discovered as a regulator of cAMP-mediated gene expression and may serve as a biomarker or a drug target in context of diseases with deregulated cAMP signaling [63]. And further investigation demonstrated that AA041523 mediated decidualization of human endometrial stromal cells and the expression of AA041523 was regulated by cAMP-PKA pathway through IL-11-mediated STAT3 phosphorylation [64]. Recently, the elevated expression of AA041523 was highly associated with loss function of the tumor suppressor LKB1 gene, one of the most common mutational events in lung cancer [65]. Further analysis suggested that AA041523 could act as a biomarker or a therapeutic target for lung cancer with impaired LKB1 signaling [65]. Additionally, AA041523 was found down-regulated in Helicobacter pylori-infected cells which might contribute to the pathological responses and development of Helicobacter pylori related disease [66]. In our study, AA041523 was also highly expressed in GC and involved with shorten survival. Thus, we suggest that AA041523 may play a critical role in GC tumorigenesis. More
importantly, BC021187 (DKFZP434K028) got a lower expression level in GC tissues and the low expression was correlated with larger tumor size [67]. In the present study, the higher level of BC021187 was associated with longer survival, suggesting a protective role in GC biogenesis. And further investigations are needed to confirm that.

Among the 24 lncRNAs, except for those mentioned above, some lncRNAs were either poorly investigated or have not been reported. For instance, BE621082 (LINC0142) was identified as a novel susceptibility locus on Xp11.21 associated with systemic lupus erythematosus (SLE) [68]. Moreover, as a super enhancer, AL703532 (CARMN) was regarded as a regulator of cardiac cell differentiation and hemeostasis [69]. Additionally, as a gene on the opposite strand of the ghrelin gene, AI056187 (GHRLOS) spanned the promoter and untranslated regions of the ghrelin gene [70] and lowly expressed in the normal gastric tissues [70]. In our study, AI056187 was highly expressed in GC and significantly correlated with longer survival which indicated a protective role in GC biogenesis. As for the rest of the lncRNAs, such as AF035291, AI028608, BC037827, BC038210, AA463827, AK056852, AW206234 and AI095542 were associated with shorten survival in our study, whereas AI080288, BF238392, BC005107, BC039674, T79746, H11436, BF511694, BC035722 were associated with prolonged survival. Although the roles of these genes in GC or other diseases biogenesis are presently unclear, our findings suggest that they deserve further investigations.

The 24-lncRNA signature is a significant determinant of survival in GC

By applying the 24-lncRNA signature to the GSE62254 test series patients, a clear separation was observed in survival curves between patients with high- and low-risk signatures. Patients with a high-risk 24-lncRNA signature in their tumor specimens tended to have shortened survival, whereas patients with a low-risk 24-lncRNA signature tended to have prolonged survival. The association between the lncRNA signature and survival was significant no matter whether the former was evaluated as a continuous variable or category variable (divided by the median cutoff). The usefulness of this lncRNA signature could be internally validated in the non-overlapping GSE62254 patients (the validation series) and another independent cohort of GSE15459 that profiled through the same platform of GSE62254, indicating good reproducibility of this 24-lncRNA signature in GC. Taken

![Fig. 6](image-url)
together, our results suggest that the 24-lncRNA signature may be a significant determinant of survival in GC, rather than an accidental feature of the transcription noise.

The 24-lncRNA signature is an independent prognostic factors in GC

Further analysis uncovered that the prognostic value of the 24-lncRNA signature was independent of one of the main prognostic factors in GC, LNR. LNR was defined as the ratio of the number of metastatic lymph nodes to the number of removed lymph nodes [47]. Recently, LNR has gained increasing attention in researches because of its lymph node status in AJCC TNM stage system [71, 72]. In Japan, LNR has been repeatedly reported to be of prognostic relevance in advanced GC in the multivariate analysis [73]. Two studies have indicated that LNR = 16.7 % was optimal cutoff level as an effective prognostic indictor in advanced GC [46, 47]. Patients with LNR ≥ 16.7 % got a shortened survival than those of LNR < 16.7 % [46, 47]. Therefore, it is important to evaluate whether the prognostic value observed on our 24-lncRNA signature is independent of this known strong prognostic factor or not. By performing multivariable Cox regression analysis and LNR stratification analysis, we identified LNR-independent prognostic values of the 24-lncRNA signature in GC patients of the entire GSE62254 data set.

Moreover, the prognostic value of the 24-lncRNA signature was also independent of the postoperative chemotherapy. Currently, surgery followed by chemoradiotherapy is the standard protocol in the United States, whereas perioperative or postoperative chemotherapy is recommended in the Europe and Asia. Increasing meta-analyses published have demonstrated that postoperative chemotherapy could prolong the survival [74, 75]. In the present study, our results indicated that patients with different prognoses could be divided into high- or low-risk group by the 24-lncRNA signature despite of the same postoperative chemotherapy stratum. And this further strongly demonstrated that the 24-lncRNA signature could act as an independent prognostic factor for GC. Finally, it was fascinating to find that the 24-lncRNA signature was almost independent of each TNM stage of AJCC and had a similar survival predictive ability as AJCC stage. Moreover, 24-lncRNA signature combined with AJCC stage had a stronger power for DFS prediction in the ROC analysis. At last, the 24-lncRNA signature may have a better predictive ability than both LNR and postoperative chemotherapy alone. Thus, the ability of our 24-lncRNA signature in identifying subgroups of GC patients with identical AJCC stage implies that the lncRNA signature may be used to refining the current prognostic model and facilitating further stratification of patients in the future clinical trials.

The implication of the study

The function of lncRNAs were more likely to correlate with their transcript abundance as they do not encode proteins [76]. Actually, lncRNAs have been demonstrated to have higher specificity than protein-coding mRNAs [77, 78]. Our findings may have clinical implications in the development of a novel, independent prognostic factor of GC. Additionally, given the expression of lncRNAs could be handled with transgene approaches such as the lncRNA interference (RNAi)-mediated gene silencing technology, for instance, knock-down of the classical lncRNA HOTAIR using specific siRNAs was indicated to be associated with the metastatic potential of breast cancer cells [54]. Although some of the 24 lncRNAs have not investigated or reported in GC or other diseases, we also have reason to believe that these lncRNAs may contribute to GC biogenesis. Of course, amount of analyses are greatly needed in the future investigations.
The limitations of the study

The limitations should be acknowledged for our study. First, since the two GEO data sets involved in this study were profiled through Affymetrix Human Genome U133 Plus 2.0 chips which represents part but not all of the possible lncRNA presents, the lncRNAs candidates identified here may not represent the complete lncRNA populations underlying GC biological behavior. Second, the DFS was regarded as the primary endpoint in the test and internal validation data set (GSE62254). Unfortunately, we could only use OS as the endpoint in external validation data set (GSE15459) because this data set did not contain the information about DFS. Despite this drawback, however, the significant and consistent correlation of the 24-lncRNA signature with OS in external validation data set indicates that it is a potential useful
prognostic marker for GC. Finally, we have no experimental data and lack information on the mechanism behind the signature IncRNAs, and experimental studies on these IncRNAs are greatly needed to provide important information to further our understanding of their functional roles in GC.

Conclusions
This study presents a powerful IncRNA signature by probing and integrating currently available microarray data. This innovative IncRNA signature showed independence of two main prognostic factors, LNR and postoperative chemotherapy. Also this IncRNA signature may contribute to personalize prediction of GC prognosis and acted as potential biomarkers for GC prognostication. The GSEA analysis suggested that this signature might involve with several cancer recurrence and metastasis associated pathways which supported the DFS predictive ability of the signature. The IncRNA profiling approach described here can also be applied in other cancers and will serve as a useful method for the systematic identification of IncRNA biomarkers in clinical practice. Future investigations will concentrate on the validation of our findings in planned clinical trials and the functional explanation of these IncRNAs.

Additional files

Additional file 1: Table S1. Clinical characteristics of 492 gastric cancer patients involved in the study. (DOCX 15 kb)

Additional file 2: Figure S1. Kaplan-Meier estimates of the disease free survival (DFS) of GEO patients using the 24-IncRNA signature, stratified by TNM stage. Entire GSE62254 set (N = 300) were first stratified by low TNM stage (I & II) and high TNM stage (III & IV). Kaplan-Meier plots were then used to visualize the survival probabilities for the high-risk versus low-risk group of patients determined on the basis of the median risk score from the GSE62254 set patients. (A) Kaplan-Meier curves for the entire GSE62254 set patients (N = 300); (B) Kaplan-Meier curves for patients with low TNM stage (N = 127); (C) Kaplan-Meier curves for patients with high TNM stage (N = 173). The tick marks on the Kaplan-Meier curves represent the censored subjects. The differences between the two curves were determined by the two-sided log-rank test. (EPS 1031 kb)

Additional file 3: Table S2. 24-IncRNA signature related signaling pathways with positive and negative enrichment score ranked by enrichment score. (XLS 267 kb)

Abbreviations
AJCC: American joint committee on cancer; AUROC: Area under receiver operating characteristic; CI: Confidence interval; DFS: Disease free survival; GC: Gastric cancer; GEO: Gene expression Omnibus; GSEA: Gene set enrichment analysis; HR: Hazard ratio; IncRNAs: Long non-coding RNAs; LNR: Lymph node ratio; OS: Overall survival; ROC: Receiver operating characteristic; UICC: Union international committee on cancer

Acknowledgements
Not applicable.

Funding
This work was supported by grants from National Natural Science Foundation of China (Grant No: 31371273).

Availability of data and materials
The gene expression data in this study can be found online at the Gene Expression Omnibus under accession numbers GSE62254 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62254) and GSE15459 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15459).

Authors' contributions
XZ, XT and CY contribute equally to the work. XZ, XT and SC drafted the manuscript. XZ, XT, CY and YT prepared all the figures and tables. All the authors reviewed and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Received: 23 June 2016 Accepted: 7 September 2016

References
1. Torre LA, Siegel RL, Jemal A. Global Cancer Incidence and Mortality Rates and Trends–An Update. Cancer Epidemiol Biomarkers Prev. 2016;25:16–27.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015;65:87–108.
3. Van Cutsem E, Sargent D, Topal B, Haustermans K, Prehen H. Gastric cancer. The Lancet. 2016. doi:10.1016/S0140-6736(16)30354-3.
4. Van Cutsem E, Dicato M, Gava R, Arber N, Bang Y, Bensol A, Cervantes A, Diaz-Rubio E, Ducrueix M, Glynn-Jones R, et al. The diagnosis and management of gastric cancer: expert discussion and recommendations from the 12th ESMO/World Congress on Gastrointestinal Cancer, Barcelona, 2010. Ann Oncol. 2011;22 Suppl 9–9.
5. Cristescu R, Lee J, Nebozyn M, Kim KM, Ting JC, Wong SS, Liu J, Yue YG, Wang J, Yu K, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med. 2015;21:449–56.
6. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol. 2010;17:1471–4.
7. Marano M, Boccard V, Bracchi B, Espisito G, Grasia M, Petriolo M, Pizzella M, Porfiola R, Reda G, Romano A, et al. Comparison of the 6th and 7th editions of the AJCC/UICC TNM staging system for gastric cancer focusing on the “N” parameter-related survival: the noninstitutional NodUs Italian study. World J Surg Oncol. 2015;13:215.
8. Milhami M, Ilmon E, Idz UO, Amend A, Demir U. Recent developments and innovations in gastric cancer. World J Gastrointest Oncol. 2012;4:307–20.
9. Tsai M-M, Wang C-S, Tsai C-Y, Chi H-C, Tseng Y-H, Lin K-H. Potential prognostic, diagnostic and therapeutic markers for human gastric cancer. World J Gastrointest Oncol. 2014;6:1379–803.
10. Ellimova E, Wadhwa R, Shizochak H, Sudo K, Estrella JS, Badgwell BD, Das P, Matamoros A, Song S, Amani JA. Molecular Biomarkers in Gastric Cancer. J Natl Comp Canc Netw. 2015;13:19–29.
11. Gibb EA, Brown CJ, Lam WL. The functional role of long non-coding RNA in human carcinomas. Mol Cancer. 2011;10:38.
12. Prensner JR, Chinnaian AM. The emergence of IncRNAs in cancer biology. Cancer Discov. 2011;1:391–407.
13. Mitra SA, Mitra AP, Triche TJ. A central role for long non-coding RNA in cancer. Front Genet. 2012;3:17.
14. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet. 2009;10:155–9.
15. Lipovich L, Johnson R, Lin CY. MacaronRNA underdogs in a microRNA world: evolutionary, regulatory, and biomedical significance of mammalian long non-protein-coding RNA. Biochim Biophys Acta. 1799;2010:597–615.
16. Huang KC, Yao P, Lau NC, Heald E, Ng SK, Brown C, Mok SC, Berkwitz RS, Ng SW. Relationship of XIST expression and responses of ovarian cancer to chemotherapy. Mol Cancer Ther. 2002;1:769–76.
17. Zhang X, Rice K, Wang Y, Chen W, Zhong Y, Nakayama Y, Zhou Y, Klubianski A. Maternally Expressed Gene 3 (MEG3) Noncoding Ribonucleic Acid Isoform Structure, Expression, and Functions. Endocrinology. 2010;151:939–47.
18. Wang Z, Jin Y, Ren H, Ma X, Wang B, Wang Y. Downregulation of the long non-coding RNA TUSC7 promotes NSCLC cell proliferation and correlates with poor prognosis. Am J Transl Res. 2016;8:680–7.

19. Qi P, Du X. The long non-coding RNAs, a new cancer diagnostic and therapeutic gold mine. Mod Pathol. 2013;26:155–65.

20. Crea F, Watashiki A, Quaglia L, Xue H, Pikor L, Panolia A, Wang Y, Lin D, Lam WL, Farrar WL, et al. Identification of a long non-coding RNA as a novel biomarker and potential therapeutic target for metastatic prostate cancer. Oncotarget. 2014;5:764–74.

21. Su X, Zhan P, Gavin PR, Morgan S, Womack C, N X, Shen D, Bang YJ, Im SA, Ho Kim W, et al. FGF2R amplification has prognostic significance in gastric cancer: results from a large international multicentre study. Br J Cancer. 2014;110:677–7.

22. Gavalos C, Jimeno A. HER2 in gastric cancer: a new prognostic factor and a novel therapeutic target. Ann Oncol. 2008;19:1523–9.

23. Lieto E, Ferraraccio F, Orditura M, Castellano P, Mura AL, Pinto M, Zamboli A, De Vita F, Galizia G. Expression of vascular endothelial growth factor (VEGF) and epidermal growth factor receptor (EGFR) is an independent prognostic indicator of worse outcome in gastric cancer patients. Ann Surg Oncol. 2008;15:69–79.

24. Lee J, Seo JW, Jun HJ, RS CS, Park SH, Park YS, Lim HY, Choi MG, Bae JH, Sohn TS, et al. Impact of MET amplification on gastric cancer possible roles as a novel prognostic marker and a potential therapeutic target. Oncol Rep. 2011;25:1517–24.

25. Yang Z, Guo X, Li G, Shi Y, Li L. Long non-coding RNAs as potential biomarkers in gastric cancer: Opportunities and challenges. Cancer Lett. 2016;371:62–70.

26. Nie PQ, Ma SY, Xie M, LiuYW, De W, Liu XH. Decreased long noncoding RNA MIR31HG is correlated with poor prognosis and contributes to cell proliferation in gastric cancer. Tumour Biol. 2016;37:7693–701.

27. Ye H, Liu K, Qian K. Overexpression of long non-coding RNA HOTTIP promotes tumor invasion and predicts poor prognosis in gastric cancer. Onco Targets Ther. 2016;9:2081–8.

28. Li R, Qian J, Wang YY, Zhang JX, You YP. Long non-coding RNA profiles reveal three subtypes in glioma. CNS Neurosci Ther. 2014;20:339–43.

29. Zhang XQ, Sun S, Lam KF, Kiang KM, Pu JK, Ho AS, Liu WM, Yue H, Fong CW, Wong TS, Leung GK. A long non-coding RNA signature in glioblastoma multiforme predicts survival. Neurobiol Dis. 2013;58:123–31.

30. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 2003;31:e15.

31. Yang F, Zhang L, Huo XS, Yuan JH, Xu D, Yuan SX, Zhu N, Zhou WP, Yang GS, Wang YZ, et al. Long non-coding RNA high expression in hepatocellular carcinoma facilitates tumor growth through enhancer of zeste homolog 2 (EZH2). Hepatology. 2011;54:1679–87.

32. Crea F, Watahiki A, Quaglia L, Xue H, Pikor L, Panolia A, Wang Y, Lin D, Lam WL, Farrar WL, et al. Identification of a long non-coding RNA as a novel biomarker and potential therapeutic target for metastatic prostate cancer. Oncotarget. 2014;5:764–74.

33. Gravalos C, Jimeno A. HER2 in gastric cancer: a new prognostic factor and a novel therapeutic target. Ann Oncol. 2008;19:1523–9.

34. Lieto E, Ferraraccio F, Orditura M, Castellano P, Mura AL, Pinto M, Zamboli A, De Vita F, Galizia G. Expression of vascular endothelial growth factor (VEGF) and epidermal growth factor receptor (EGFR) is an independent prognostic indicator of worse outcome in gastric cancer patients. Ann Surg Oncol. 2008;15:69–79.

35. Lee J, Seo JW, Jun HJ, RS CS, Park SH, Park YS, Lim HY, Choi MG, Bae JH, Sohn TS, et al. Impact of MET amplification on gastric cancer possible roles as a novel prognostic marker and a potential therapeutic target. Oncol Rep. 2011;25:1517–24.

36. Yang Z, Guo X, Li G, Shi Y, Li L. Long non-coding RNAs as potential biomarkers in gastric cancer: Opportunities and challenges. Cancer Lett. 2016;371:62–70.

37. Nie PQ, Ma SY, Xie M, LiuYW, De W, Liu XH. Decreased long noncoding RNA MIR31HG is correlated with poor prognosis and contributes to cell proliferation in gastric cancer. Tumour Biol. 2016;37:7693–701.

38. Ye H, Liu K, Qian K. Overexpression of long non-coding RNA HOTTIP promotes tumor invasion and predicts poor prognosis in gastric cancer. Onco Targets Ther. 2016;9:2081–8.

39. Li R, Qian J, Wang YY, Zhang JX, You YP. Long non-coding RNA profiles reveal three subtypes in glioma. CNS Neurosci Ther. 2014;20:339–43.

40. Zhang XQ, Sun S, Lam KF, Kiang KM, Pu JK, Ho AS, Liu WM, Yue H, Fong CW, Wong TS, Leung GK. A long non-coding RNA signature in glioblastoma multiforme predicts survival. Neurobiol Dis. 2013;58:123–31.

41. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 2003;31:e15.

42. Yang F, Zhang L, Huo XS, Yuan JH, Xu D, Yuan SX, Zhu N, Zhou WP, Yang GS, Wang YZ, et al. Long non-coding RNA high expression in hepatocellular carcinoma facilitates tumor growth through enhancer of zeste homolog 2 (EZH2). Hepatology. 2011;54:1679–87.

43. Huo XS, Yuan JH, Xu D, Yuan SX, Zhu N, Zhou WP, Yang GS, Wang YZ, et al. Long non-coding RNA high expression in hepatocellular carcinoma facilitates tumor growth through enhancer of zeste homolog 2 (EZH2). Hepatology. 2011;54:1679–87.

44. Crea F, Watahiki A, Quaglia L, Xue H, Pikor L, Panolia A, Wang Y, Lin D, Lam WL, Farrar WL, et al. Identification of a long non-coding RNA as a novel biomarker and potential therapeutic target for metastatic prostate cancer. Oncotarget. 2014;5:764–74.

45. Gravalos C, Jimeno A. HER2 in gastric cancer: a new prognostic factor and a novel therapeutic target. Ann Oncol. 2008;19:1523–9.
66. Zhu H, Wang Q, Yao Y, Fang J, Sun F, Ni Y, Shen Y, Wang H, Shao S. Microarray analysis of Long non-coding RNA expression profiles in human gastric cells and tissues with Helicobacter pylori Infection. BMC Med Genomics. 2015;8:84.

67. Zhao J, Liu Y, Zhang W, Zhou Z, Wu J, Cui P, Zhang Y, Huang G. Long non-coding RNA Linc00152 is involved in cell cycle arrest, apoptosis, epithelial to mesenchymal transition, cell migration and invasion in gastric cancer. Cell Cycle. 2015;14:3112–23.

68. Zhu Z, Liang Z, Liang Y, Yang C, Wen L, Lin Z, Sheng D, Liu Y, Ye L, Cheng Y, et al. Discovery of a novel genetic susceptibility locus on X chromosome for systemic lupus erythematosus. Arthritis Res Ther. 2015;17:349.

69. Ounzain S, Micheletti R, Arnan C, Plaisance I, Cecchi D, Schroen B, Reverter F, Alexanian M, Gonzales C, Ng SY, et al. CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeostasis. J Mol Cell Cardiol. 2015;89:998–112.

70. Seim I, Carter SL, Herington AC, Chopin LK. Complex organisation and structure of the ghrelin antisense strand gene GHRLOS, a candidate non-coding RNA gene. BMC Mol Biol. 2008;9:95.

71. Wang J, Dang P, Raut CP, Pandalai PK, Maduwwe UN, Rattner DW, Lauwers GY, Yoon SS. Comparison of a lymph node ratio-based staging system with the 7th AJCC system for gastric cancer: analysis of 18,043 patients from the SEER database. Ann Surg. 2012;255:478–85.

72. Lee SR, Kim HO, Son BH, Shin JH, Yoo CH. Prognostic significance of the metastatic lymph node ratio in patients with gastric cancer. World J Surg. 2012;36:1096–101.

73. Fukuda N, Sugiyama Y, Midorikawa A, Mushiake H. Prognostic significance of the metastatic lymph node ratio in gastric cancer patients. World J Surg. 2009;33:2378–82.

74. Panzini I, Gianni L, Fattori PP, Tassinari D, Imola M, Fabbri P, Arcangeli V, Drudi G, Canuti D, Fochessati F, Ravalotti A. Adjuvant chemotherapy in gastric cancer: a meta-analysis of randomized trials and a comparison with previous meta-analyses. Tumori. 2002;88:21–7.

75. Oba K, Morita S, Tsuburaya A, Kodera Y, Kobayashi M, Sakamoto J. Efficacy of adjuvant chemotherapy using oral fluorinated pyrimidines for curatively resected gastric cancer: a meta-analysis of centrally randomized controlled clinical trials in Japan. J Chemother. 2006;18:311–7.

76. Du Z, Fei T, Verhaak RG, Su Z, Zhang Y, Brown M, Chen Y, Liu XS. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. Nat Struct Mol Biol. 2013;20:908–13.

77. Hessels D, Klein Gunnewiek JM, van Oort I, Karthaus HF, van Leenders GJ, van Balken B, van Riel M, Witjes JA, Schalken JA. DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. Eur Urol. 2003;44:148–15. discussion 15-16.

78. Prentine IR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, Laxman B, Asangani IA, Grasso CS, Kominsky HD, et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nat Biotechnol. 2011;29:742–9.