LncRNA expression signature identified using genome-wide transcriptomic profiling to predict lymph node metastasis in patients with stage T1 and T2 gastric cancer

Zhebin Dong
Ningbo Medical Centre Lihuili Hospital
Hanting Xiang
Ningbo Medical Centre Lihuili Hospital
Hengmiao Wu
Ningbo Medical Centre Lihuili Hospital
Zhengwei Chen
Ningbo Medical Centre Lihuili Hospital
Sangsang Chen
Ningbo Medical Centre Lihuili Hospital
Yicheng He
Ningbo Medical Centre Lihuili Hospital
Hong Li
Ningbo Medical Centre Lihuili Hospital
Weiming Yu
Ningbo Medical Centre Lihuili Hospital
Chao Liang (movingstar-lchao@163.com)
Ningbo Medical Centre Lihuili Hospital

Research Article

Keywords: Bioinformatics, Biomarkers, Gastric cancer, lymph node, Metastasis, Transcriptomic, LncRNA signature

Posted Date: March 1st, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2618696/v1

Abstract

Background: Lymph node (LN) status is vital to indicate and evaluate the curative potential of relatively early gastric cancer (GC; T1–T2) treatment (endoscopic or surgery). Currently, there is a lack of robust and convenient methods to identify such metastasis before therapeutic decision-making; therefore, there is an urgent need to identify biomarkers that could aid the identification of patients with LN metastasis.

Methods: Genome-wide expression profiles of long noncoding RNA (lncRNA) in primary T1 gastric cancer data from The Cancer Genome Atlas (TCGA) was used to identify an lncRNA-expression signature capable of detecting LN metastasis of GC, and establish a 10-lncRNA risk-prediction model based on deap learning. The performance of the lncRNA panel in diagnosing LN metastasis was evaluated using both in silico and clinical validation methods. In silico validation was conducted using TCGA and Asian Cancer Research Group (ACRG) datasets. Clinical validation was performed on T1 and T2 patients, and the panel's efficacy was compared with that of traditional tumor markers and computed tomography (CT) scans.

Results: Profiling of genome-wide RNA expression identified a panel of lncRNA to predict LN metastasis in T1 stage gastric cancer (area under the curve (AUC) = 0.961). A 10-lncRNA risk-prediction model was then constructed, which was validated successfully in T1 and T2 datasets (TCGA, AUC = 0.852; ACRG, AUC = 0.834). Thereafter, the clinical performance of the lncRNA panel was validated in clinical cohorts (T1, AUC = 0.812; T2, AUC = 0.805; T1+T2, AUC = 0.764). Notably, the 10-lncRNA panel demonstrated significantly better performance compared with CT and conventional tumor markers (carcinoembryonic antigen and carbohydrate antigen 19-9).

Conclusions: The novel 10-lncRNA could diagnose LN metastasis robustly in relatively early gastric cancer (T1–T2), with promising clinical potential.

Background

Globally, gastric cancer (GC) is fifth in the list of most common cancers and ranks fourth as the most common cause of cancer death[1]. Lymph node (LN) metastasis is a major clinical feature of GC, which influences the poor prognosis of patients with GC [2]. Even for early GC, the 10-year survival rates of patients with or without LN metastasis are significantly different, at 72 and 92%, respectively [3]. Accurate evaluation of LN status in patients with GC before treatment is critical to evaluate the degree of disease and improve treatment strategies. Currently, the diagnosis of LN metastasis is carried out mainly using conventional tumor markers (carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9)) and computed imaging methods (computed tomography (CT) and positron emission tomography with CT (PET-CT)). Unfortunately, these methods show poor performance to clinically identify LN and frequently demonstrate poor correlation and high error rates [4, 5]. Thus, there is an urgent need for more accurate and reliable detection methods to identify LN metastasis in GC, which might be used to enhance the prognosis of patients with GC significantly.

Gastric cancer is still treated using surgery and endoscopic resection. Currently, Asian and European guidelines identify endoscopic submucosal dissection (ESD) and endoscopic mucosal resection (EMR) as the first choice treatments for most cases of early GC (cT1a) and are considered to be safe and definitive treatments[6, 7]. However, patients who are considered to be at risk of LN metastasis after endoscopic surgery will undergo additional radical surgery, because of submucosal invasion (T1b), large tumor size, and poor differentiation [8]. Unfortunately, pathological examination of these post-gastrectomy tissues, especially from early GC, revealed that only about 20% of patients were identified as having LN metastasis [9, 10]. In past decades, the optimal extent of lymphadenectomy has also been discussed extensively in the field of surgery. With the development of precision medicine, for patients with GC with cT1-T2N0M0 status, laparoscopic sentinel node navigation surgery (LSNNS) was proposed for stomach preservation, which showed no difference in three year overall survival (3y-OS) and three year disease free survival (3y-DFS) compared with laparoscopic standard D2 gastrectomy, but resulted in better long-term quality of life and nutritional status [11, 12]. Prospective evaluation of sentinel lymph node navigation surgery for relatively early GC (T1–T2) is a current development trend of function-preserving, personalized, and minimized gastrectomy [12-14]. However, LSNNS is based on a comprehensive assessment of the LN status of patients, which is a challenge for its practical application. The lack of accurate and reliable detection of preoperative LN metastasis status means that many patients have experienced unnecessary overtreatment, which also limits the beneficial development of precision medicine.

Herein, transcriptome-wide expression profiles of long non-coding RNA (lncRNA) were analyzed comprehensively and systematically, and a 10-lncRNA panel was established to identify GC LN metastasis (T1 and T2). We verified the effectiveness of the panel in independent databases and clinical tissue samples. The performance of the lncRNA panel was also compared with that of CEA, CA19-9, and CT, highlighting the value of this panel in predicting LN metastasis of T1 and T2 GC. The lncRNA panel could function as the basis for clinical decision-making for patients with GC

Methods

Public Datasets and the Identification of the Gene-expression Signature.

To identify an lncRNA expression signature for the detection of lymph node (LN) metastasis in gastric cancer (GC). The study used genome-wide expression profiles of lncRNAs from primary tumors with and without LN metastasis, which were obtained from The Cancer Genome Atlas (TCGA) database. Only pathological T1 and T2 RNA-sequencing (RNA-Seq) data were used for further analysis. The T1 data was from 15 LN metastasis negative (LNN) and 5 LN metastasis positive (LNP) samples, and the T2 data was from 34 LNN and 48 LNP samples. The processed TCGA level 3 RNA-Seq data for GC was obtained from the Firehose Broad GDAC portal[15]. Independent validation data were downloaded from the Asian Cancer Research Group (ACRG). In the gene-level RNA-Seq by Expectation-Maximization (RSEM files), we converted the scaled estimates to transcripts per million (TPM) by multiplying them by 10^6, and then carrying out log2-transformation. We filtered all lncRNA expression levels from the TCGA and ACRG

processed data according to the human gene annotation file [16] (https://ftp.ensembl.org/pub/release97/gtf/homo_sapiens/Homo_sapiens.GRCh38.97.chr.gtf.gz). Then, logistic regression analysis was performed using the Logistic Regression (LR) function from Pytorch [17] (citation https://arxiv.org/abs/1912.01703). Feature importance was estimated using coefficients from the LR model. To assess the lncRNA panel's diagnostic accuracy, the selected lncRNA features were used to construct a multivariate LR model, followed by calculation of the area under the receiver operator characteristic (ROC) curve (AUC) values. Ultimately, the probability of each patient being identified as LNP was used as the basis to calculate the risk scores.

Clinical cohort evaluation

To validate the identified lncRNA markers and for clinical training, we enrolled three independent patient cohorts comprising 245 cases in total. Cohort-1 consisted of 20 surgically resected GC specimens from 8 LNP patients and 12 LNN patients. Cohort-2 included 98 patients (LNP = 19, LNN = 79). Cohort-3 included 127 patients (LNP = 38, LNN = 89). Patients in the clinical cohorts were treated at the Lihuili Hospital affiliated to Ningbo University (China). These patients had biopsy-proven primary GC and underwent curative surgery between December 2017 and January 2022. During surgery, we obtained tissues samples from a representative malignant lesion located in the surgically excised stomach specimen. The tissue samples were added with RNAstore (CWBIO, Shanghai, China), frozen rapidly in liquid nitrogen, and stored at -80 ºC. The summarized characteristics of the patients in the clinical cohorts are shown in Table 1 and Fig. 2A.

RNA Extraction and Quantitative Real-time Reverse Transcription PCR (qRT-PCR) Analysis

An RNeasy mini kit (Qiagen, Hilden, Germany) was used to isolate total RNA from frozen surgical tissues, following the supplier's guidelines. The RT-PCR step of the qRT-PCR protocol was carried out using a SensiFAST probe Lo-ROX Kit (Bioline, London, UK) and the qPCR step used the QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). Assay reproducibility was ensured via multiple techniques, such as including appropriate controls, excluding specimens with poor RNA quality, and the analysis of multiple replicates carried out at various time points. The QuantStudio 6 Flex Real-Time PCR System Software (Applied Biosystems) was used to assess gene expression. The expression level of *ACTB* (encoding beta actin) was used to determine and correct the relative expression of target genes, employing the $2^{-\Delta Ct}$ method. In this method, ΔCt is the difference in cycle threshold (C_t) values between *ACTB* and the gene of interest. The data were then $\log_2$ transformed. Supplementary Table S1 details the PCR primers used.

Statistical Evaluations

The method of *DeLong* [18] was used to assess the statistically significant differences among the ROC curves. Python (version 3.8, https://www.python.org/) was used to carry out the statistical analyses. Two-tailed t test-determined p values less than 0.05 indicated statistical significance.

LncRNA enrichment analysis

For lncRNA enrichment analysis, we used the website application constructed by Chen et al[19]. (https://doi.org/10.1093/nar/gkaa806). The data were visualized on a histogram and bubble chart using ggplot2 [20].

Results

Genome-wide lncRNA expression profiles identified a 10-lncRNA panel to predict LN metastasis in T1 and T2 stage GC

First, we systematically and comprehensively analyzed RNA-seq expression profiling data from patients with GC at T1 stage in the TCGA database, which included 5 LNP patients and 15 LNN patients, to identify an lncRNA expression signature to diagnose patients with T1 stage GC with LN metastasis using deep learning model. The validation of the holdout dataset demonstrated that the model could distinguish patients with LNP GC from those with LNN GC (AUC = 0.961, Figure 1A). In order to make the lncRNA signature more practical and suitable for clinical use, we prioritized lncRNAs resulted in a10-lncRNA signature for further validation based on the feature importance in logisctic regression, which included five relatively highly expressed lncRNAs (*H19*, *CECR7*, *HOTAIR*, *FAM66D*, *C22orf34*) and five lncRNAs with relatively low expression (*TTTY15*, *TTTY14*, *TP53TG1*, *HAR1A*, *C10orf95*) in the LNP versus LNN comparison. In addition, tumor functional enrichment analysis was carried out for this panel. The results revealed that this 10-lncRNA panel was closely related to tumor prognosis, epithelial-mesenchyme transition (EMT), and metastasis, and was specific for gastrointestinal system cancer (Figure 1B).

Considering that both T1N0M0 and T2N0M0 are stage I gastric cancer, and the significance of T2 lesions in the current precision medicine of GC, the predictive accuracy of the 10-lncRNA panel was also validated in the dataset containing patients with T2 GC. Notably, individual lncRNAs showed limited performance in external independent datasets, while the integration of all 10 lncRNAs demonstrated significant performance. Multivariate LR analysis was then used to obtain a 10-lncRNA risk-prediction model: risk score = -0.141 * *TTTY15* - 0.140 * *TTTY14* - 0.117 * *TP53TG1* - 0.100 * *HAR1A* - 0.074 * *C10orf95* + 0.166 * *H19* + 0.212 * *CECR7* + 0.222 * *HOTAIR* + 0.226 * *FAM66D* + 0.236 * *C220rf34*. Using a larger TCGA cohort (including 20 T1 and 82 T2 patients) and a cohort of 186 T2 patients from the ACRG (Figure 1C,D), the risk model could differentiate LNP from LNN patients (AUC = 0.85, Figure. 1C; AUC = 0.83, Figure 1D). This 10-lncRNA panel exhibited a robust performance in two independent validated datasets, highlighting its potential for diagnostic prediction of LN metastasis in patients with T1 and T2 stage GC.

Validation of the 10-lncRNA risk-prediction model to identify lymph node metastasis in independent clinical cohorts

The accuracy of diagnosis using the 10-lncRNA panel was assessed using RNA-seq in validation clinical cohort 1 and by qRT-PCR in validation clinical cohorts 2 and 3 (Table 1). All patients in cohort 1 were in

T1 stage, which included 8 LNP and 12 LNN patients. The heatmap of the 10-lncRNA panel and the risk score curve are shown in Figure2A. As expected, there was a significant difference in the expression of corresponding lncRNAs between the LNN and LNP samples, revealing an effective diagnostic performance by our risk-prediction model (Figure 2A)

The patients in cohort 2 and 3 had T1 (19 LNP and 79 LNN) stage and T2 (38 LNP and 89 LNN) stage GC, respectively. Multivariate LR analysis was used to assess the effectiveness of the 10-lncRNA panel in T1–T2 tumors. In cohort 2, the panel showed an AUC of 0.812 (Table 2 and Figure 2B), and in cohort 3, the AUC was 0.805 (Table 2 and Figure 2C). However, the predictive performance of our panel in cohort 2+3 (T1+T2) was slightly reduced (AUC = 0.764, Figure 2D) compared with verification using the individual cohorts, although there was still a good effect. This might be related to the high heterogeneity of GC and the difference in the overall expression of 10 lncRNAs in the T1 and T2 specimens. Overall, the validation results agreed with those obtained using the training cohort: the 10-lncRNA panel could robustly and effectively distinguish LNP from LNN in patients with T1 and T2 stage GC.

Table 1. Clinical characteristics of the patients in cohort 2 and 3

characteristics	Clinical cohort-2 (n = 98)		Clinical cohort-3 (n = 127)	
	LN positive (n = 19)	LN negative (n = 79)	LN positive (n = 38)	LN negative (n = 89)
Age (years)	69.53 ± 6.141	68.62 ± 7.550	69.42 ± 8.465	69.00 ± 0.769
Sex				
Male	11	40	23	49
Female	8	39	15	40
CEA (ng/ml)	4.226±3.046	2.890±1.597	4.750±3.307	3.794±2.098
positive	7	8	10	21
negative	12	71	28	68
CA19-9 (U/ml)				
positive	8	8	9	14
negative	11	71	29	75
CT				
positive	5	9	12	10
negative	14	70	26	79
T stage				
T 1	19	79	0	0
T 2	0	0	38	89
N stage				
N 0	0	79	0	89
N 1	16	0	25	0
N 2	3	0	10	0
N 3	0	0	3	0
LV invasion				
positive	15	8	30	35
negative	4	71	8	54
Venous invasion				
positive	12	15	26	34

negative	7	64	12	55

The CEA cutoff value is 5 ng/ml; the CA19-9 cutoff value is 37 U/ml. LN, lymph node; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; CT computed tomography; T stage, tumor stage; N stage, node stage; LV, lymphovascular.

Table 2. Summary of the individual lncRNA performance to predict lymph node metastasis in clinical cohorts 2 and 3

lncRNA	Clinical cohort-2			Clinical cohort-3		
	AUC	Specificity	sensitivity	AUC	Specificity	sensitivity
TTTY15	0.659	0.947	0.557	0.586	0.658	0.584
TTTY14	0.526	0.368	0.949	0.656	0.579	0.753
TP53TG1	0.610	0.895	0.443	0.644	0.447	0.865
HAR1A	0.562	0.684	0.544	0.670	0.526	0.798
C10orf95	0.672	0.737	0.595	0.690	0.526	0.888
H19	0.676	0.772	0.526	0.676	0.629	0.684
CECR7	0.709	0.709	0.789	0.669	0.742	0.737
HOTAIR	0.710	0.975	0.421	0.689	0.921	0.500
FAM66D	0.701	0.709	0.842	0.707	0.843	0.632
C22orf34	0.702	0.570	0.947	0.695	0.730	0.711
Risck Score	0.812	0.861	0.684	0.805	0.753	0.816

AUC, area under the curve.

The 10-lncRNA panel showed better diagnostic power compared with conventional tumor markers and CT in LN metastasis.

For the surgical management of patients with GC, generally, enhanced CT imaging is employed to determine clinical N stage before surgery. Typically, CT features such as an LN diameter ≥ 1cm, ring or heterogeneous enhancement, are employed to diagnose LN metastasis. However, CT imaging cannot successfully diagnose most cases of LN metastasis, or there may be misdiagnosis; therefore, only the pathological examination of surgically excised tissue can confirm LM metastasis in such cases. As shown in Table 3, we demonstrated that our 10-lncRNA risk-prediction model could effectively identify LN metastasis using univariate and multivariate analysis, independent of preoperative clinical characteristics such as sex, age, conventional tumor markers, and CT.

To assess the diagnostic efficiency of the panel, its performance was compared with that of conventional tumor markers (CEA and CA19-9) and CT in clinical cohort. Our 10-lncRNA panel showed significant superiority over preoperative clinical factors, CEA, CA19-9, and CT (Figure 2B, 2C and 2D, comparison of the AUC values were compared using the *DeLong* test). In addition, we combined the 10-lncRNA panel with clinicopathological features (CEA, CA19-9, and CT) in cohorts 2+3. The results was also encouraging: this combination further improved the diagnostic accuracy of our panel (AUC = 0.813) compared with the 10-lncRNA panel alone (Figure 2E). In conclusion, we constructed and validated a 10-lncRNA panel that demonstrated robust discriminative power compared with current preoperative management approaches to identify cases of LNP gastric cancer.

Table 3. Univariate and multivariate logistic regression analysis of the statistical significance of the 10-lncRNA risk score to diagnosing LN metastasis status in clinical cohorts 2 and 3

	Univariate analysis			Multivariate analysis		
Variables	Odds Ratio	95% CI	*P*	Odds Ratio	95% CI	*P*
Age	1.012	0.971-1.054	0.578	1.004	0.955-1.057	0.888
Sex	1.312	0.713-2.414	0.383	1.421	0.687-2.939	0.342
CEA	2.037	1.017-4.079	0.045	1.528	0.654-3.572	0.327
CA19-9	2.820	1.369-5.813	0.005	4.104	1.677-10.046	0.002
CT	3.333	1.588-6.997	0.001	4.073	1.661-9.991	0.002
Risk score	8.889	4.085-19.341	< 0.0001	11.072	4.706-26.046	< 0.0001

LN, lymph node; CI, confidence interval; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; CT,computed tomography.

Discussion

Currently, minimally invasive or non-invasive, stomach-preserving, function-preserving, and individualized treatment has become a trend in global GC treatment. Clinically, determining LN status is crucial to indicate and evaluate the curative potential of GC endoscopic treatment and surgery, especially in patients with relatively early GC (T1–T2). Pathological diagnosis following radical gastrectomy remains the optimal way to evaluate a patient's GC's LN status, considering our lack of effective molecular markers that can robustly detect LN metastasis before therapeutic decision-making. Moreover, only patients with GC *in situ* (Tis stage) and T1a GC without LN metastasis can be treated successfully using endoscopic mucosal or submucosal resection. However, the actual LN metastasis rate of early GC (T1) is only around 20 %. In addition, the incidence of regional LN metastasis is limited in patients with T2 GC, in which D2 gastrectomy might be an excessively invasive surgery, involving in a significant waste of medical resources [9, 10, 12]. Currently, the development of sentinel node navigation surgery (SNNS) and

laparoscopic surgery in GC provides a direction for minimally invasive gastric surgery. The study group of the Japan Society of SNNS has already formulated the standard procedure for SNNS, which uses a dual tracer comprising technetium 99m–labeled tin colloid and 1% isosulfan blue dye [21]. Although several single institutions have reported the successful use of SNNS, because GC has a somewhat complex lymphatic flow, there still are controversial aspects regarding the application of SNNS [12, 22, 23].

LncRNAs are mRNA-like transcripts of >200 nucleotides with no capacity to encode proteins (21). A variety of cancers show abnormal expression of lncRNAs, which have diverse functions in gene regulation, cell biological behavior, and tumor initiation and progression [25, 26]. To date, there have been a considerable number of studies on lymph node metastasis of GC; however, most of them explored the regulatory mechanism of a single lncRNA [27, 28]. Although these studies are meaningful and significant, the lack of a comprehensive and dynamic understanding of lymph node metastasis limits the clinical application value of these findings. The recent development and popularization of high-throughput sequencing technologies have increased our understanding of the molecular characteristics of GC [29, 30]. Notably, the different T stages of GC have strong histological heterogeneity, and the correlation between lncRNAs and LN metastasis in relatively early GC (T1-T2) remains unexplored.

In this article, we used RNA sequencing to gain insights into the molecular biology of tumor heterogeneity and disease processes to identify LN metastasis. A systematic and comprehensive analysis of transcriptome-wide expression profiles of patients with T1––T2 GC, with and without LN metastasis, was used to establish an optimized 10-lncRNA panel to identify LN metastasis using logistic regression analysis. Subsequently, the panel was validated in three independent validation cohorts based on RNA-seq and qRT-PCR, achieving encouraging results. In addition, further analysis demonstrated the superiority of the 10-lncRNA panel over current clinicopathological factors, including CEA, CA19-9, and CT-based imaging, to diagnose LN metastasis in patients with GC. Although the accuracy of 10-lncRNA panel in combined cohort 2+3 was slightly decreased, its diagnostic accuracy improved again after combining it with clinicopathological features. We also performed functional and expression enrichment analysis of the 10 lncRNAs, several of which are related to metastasis and prognosis. LncRNA *H19* is considered a carcinogenic factor in GC, and its upregulation is related to tumor cell proliferation, invasion, migration, and EMT [31]. *HOTAIR* has been reported to be related to the expression of *HER2* (encoding human epidermal growth factor receptor 2) and facilitates GC lymph node metastasis [32]. A study using TCGA-based bioinformatics analysis and microarray analysis revealed that *HAR1A* is a tumor suppressor involved in tumor progression via EMT regulation and is negatively associated with prognosis [33, 34]. In our panel, *HAR1A* also acted as a negative factor for early lymph node metastasis in GC. Similarly, *TP53TG1* and *TTTY15* have been confirmed to be differentially expressed in GC tissues compared with that in normal gastric mucosa [25, 35]. Finally, as biomarkers, each lncRNA in our panel was endowed with an additional diagnostic coefficient and made a significant contribution to the identification of LN metastasis.

This study has certain limitations. First, this was a retrospective study, and its design means that although we validated our findings in multiple clinical cohorts, prospective studies are still required.

Second, the main aim of this study was to find early-stage GC biomarkers; therefore, the samples were concentrated in the T1 and T2 GC stages, which limited the sample size to discover biomarkers and had a certain impact on obtaining the panel with maximum efficiency. To overcome these limitations, larger cohorts comprising patients with GC and T1 and T2 LN metastasis are required, which might involve the participation of multiple medical institutions.

Conclusion

Our panel provided and validated a class of biomarkers that could robustly categorize patients with relatively early GC according to their LN status prior to therapeutic decision-making, thus permitting individualized treatment. Our panel offers promising diagnostic potential to identify patients with GC with or without LN metastasis; however, our findings should be validated prospectively using clinical cohorts.

Abbreviations

LN, lymph node; GC, gastric cancer; lncRNA, long noncoding RNA; TCGA, The Cancer Genome Atlas; ACRG, Asian Cancer Research Group; CT, computed tomography ; AUC, area under the curve; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; PET-CT, positron emission tomography with CT; ESD, endoscopic submucosal dissection; EMR, endoscopic mucosal resection; LSNNS, laparoscopic sentinel node navigation surgery; OS, overall survival; DFS, disease free survival; RNA-Seq, RNA-sequencing; LNN, lymph node metastasis negative; LNP, lymph node metastasis positive; TPM, transcripts per million; LR, Logistic Regression; ROC, receiver operator characteristic; qRT-PCR, quantitative real-time reverse transcription PCR; EMT, epithelial-mesenchyme transition; SNNS, sentinel node navigation surgery

Declarations

Funding

This work was supported by grants from the National Natural Science Funds of Young Scientists of China (No. 81802944), the Key Project of Ningbo Natural Science Foundation (No. 2021J279), the Zhejiang Province Medical and Health Project (No. 2023KY236), the Ningbo Natural Science Foundation (No. 2022J264), Ningbo Digestive System Tumor Clinical Medicine Research Center (No.2019A21003), Project of Zhejiang Medical and Health Platform Plan (No.2022KY1079) and Ningbo Public Welfare Science & Technology Major Project (No.2021S106).

Authors' contributions

Zhebin Dong: Data curation; Project administration; Writing – original draft. Hanting Xiang: Data curation; Formal analysis; Project administration. Hengmiao Wu: Data curation; Project administration. Zhengwei Cheng: Investigation. Sangsang Chen: Investigation. Yicheng He: Investigation. Hong Li: Conceptualization; Methodology; Writing – review & editing. Weiming Yu: Conceptualization;

Methodology; Writing – review & editing. Chao Liang: Conceptualization; Methodology; Writing – review & editing.

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Affiliated Lihuili Hospital, Ningbo University (no.KY2022SL436-01) on 13 December 2022, with an exemption from informed consent. No specific consent is needed for statistical analyses of aggregated deidentified data. For this study, the raw data were first extracted from HIS, and patients' identities, including names, screening IDs, patient IDs, and mobile phone numbers, were de-identified.

Consent for publication

All authors approved the final manuscript and the submission to this journal.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

All the Authors have no conflict of interest related to the manuscript.

Acknowledgements

None.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F: **Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries**. *Ca-a Cancer Journal for Clinicians* 2021, **71**(3):209-249.
2. Brown M, Assen FP, Leithner A, Abe J, Schachner H, Asfour G, Bago-Horvath Z, Stein JV, Uhrin P, Sixt M *et al*: **Lymph node blood vessels provide exit routes for metastatic tumor cell dissemination in mice**. *Science* 2018, **359**(6382):1408-+.
3. Ren G, Cai R, Zhang W-J, Ou J-M, Jin Y-N, Li W-H: **Prediction of risk factors for lymph node metastasis in early gastric cancer**. *World Journal of Gastroenterology* 2013, **19**(20):3096-3107.
4. Feng F, Tian Y, Xu G, Liu Z, Liu S, Zheng G, Guo M, Lian X, Fan D, Zhang H: **Diagnostic and prognostic value of CEA, CA19-9, AFP and CA125 for early gastric cancer**. *Bmc Cancer* 2017, **17**.
5. Saito T, Kurokawa Y, Takiguchi S, Miyazaki Y, Takahashi T, Yamasaki M, Miyata H, Nakajima K, Mori M, Doki Y: **Accuracy of multidetector-row CT in diagnosing lymph node metastasis in patients with gastric cancer**. *European Radiology* 2015, **25**(2):368-374.

6. Japanese Gastric Canc A: **Japanese gastric cancer treatment guidelines 2014 (ver. 4)**. *Gastric Cancer* 2017, **20**(1):1-19.
7. Pimentel-Nunes P, Libanio D, Bastiaansen BAJ, Bhandari P, Bisschops R, Bourke MJ, Esposito G, Lemmers A, Maselli R, Messmann H *et al*: **Endoscopic submucosal dissection for superficial gastrointestinal lesions: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2022**. *Endoscopy* 2022, **54**(06):591-622.
8. Hatta W, Gotoda T, Koike T, Masamune A: **History and future perspectives in Japanese guidelines for endoscopic resection of early gastric cancer**. *Digestive Endoscopy* 2020, **32**(2):180-190.
9. Hirasawa T, Gotoda T, Miyata S, Kato Y, Shimoda T, Taniguchi H, Fujisaki J, Sano T, Yamaguchi T: **Incidence of lymph node metastasis and the feasibility of endoscopic resection for undifferentiated-type early gastric cancer**. *Gastric Cancer* 2009, **12**(3):148-152.
10. Gotoda T, Yanagisawa A, Sasako M, Ono H, Nakanishi Y, Shimoda T, Kato Y: **Incidence of lymph node metastasis from early gastric cancer: estimation with a large number of cases at two large centers**. *Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association* 2000, **3**(4):219-225.
11. Kim Y-W, Min J-S, Yoon HM, An JY, Eom BW, Hur H, Lee YJ, Cho GS, Park Y-K, Jung MR *et al*: **Laparoscopic Sentinel Node Navigation Surgery for Stomach Preservation in Patients With Early Gastric Cancer: A Randomized Clinical Trial**. *Journal of Clinical Oncology* 2022, **40**(21):2342-+.
12. Kitagawa Y, Takeuchi H, Takagi Y, Natsugoe S, Terashima M, Murakami N, Fujimura T, Tsujimoto H, Hayashi H, Yoshimizu N *et al*: **Sentinel Node Mapping for Gastric Cancer: A Prospective Multicenter Trial in Japan**. *Journal of Clinical Oncology* 2013, **31**(29):3704-+.
13. Eom BW, Yoon HM, Min JS, Cho I, Park J-H, Jung MR, Hur H, Kim Y-W, Park YK, Nam B-H *et al*: **Prospective Multicenter Feasibility Study of Laparoscopic Sentinel Basin Dissection after Endoscopic Submucosal Dissection for Early Gastric Cancer: SENORITA 2 Trial Protocol**. *Journal of Gastric Cancer* 2019, **19**(2):157-164.
14. Shoji Y, Kumagai K, Kamiya S, Ida S, Nunobe S, Ohashi M, Yoshimizu S, Horiuchi Y, Yoshio T, Ishiyama A *et al*: **Prospective feasibility study for single-tracer sentinel node mapping by ICG (indocyanine green) fluorescence and OSNA (one-step nucleic acid amplification) assay in laparoscopic gastric cancer surgery**. *Gastric Cancer* 2019, **22**(4):873-880.
15. The Firehose Broad GDAC portal. **http://gdac.broadinstitute.org.** Accessed 1 July 2022.
16. The human gene annotation file. **https://ftp.ensembl.org/pub/release97/gtf/homo_sasapie/Homo_sapiens.GRCh38.97.chr.gtf.gz.** Accessed 3 July 2022.
17. Pytorch. **https://arxiv.org/abs/1912.01703.** Accessed 2 July 2022.
18. DeLong ER, David M. DeLong, and Daniel L. Clarke-Pearson: **Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach**: Biometrics; 1988.
19. Chen J, Zhang J, Gao Y, Li Y, Feng C, Song C, Ning Z, Zhou X, Zhao J, Feng M, Zhang Y, Wei L, Pan Q, Jiang Y, Qian F, Han J, Yang Y, Wang Q, Li C. **LncSEA: a platform for long non-coding RNA related**

sets and enrichment analysis. Nucleic Acids Res. 2021 Jan 8;49(D1):D969-D980.

20. Hadley Wickham. **ggplot2: Elegant Graphics for Data Analysis.** Springer New York, NY; 2009.
21. Uenosono Y, Natsugoe S, Ehi K, Arigami T, Hokita S, Aikou T: **Detection of sentinel nodes and micrometastases using radioisotope navigation and immunohistochemistry in patients with gastric cancer**. *British Journal of Surgery* 2005, **92**(7):886-889.
22. Ichikura T, Sugasawa H, Sakamoto N, Yaguchi Y, Tsujimoto H, Ono S: **Limited Gastrectomy With Dissection of Sentinel Node Stations for Early Gastric Cancer With Negative Sentinel Node Biopsy**. *Annals of Surgery* 2009, **249**(6):942-947.
23. Yano K, Nimura H, Mitsumori N, Takahashi N, Kashiwagi H, Yanaga K: **The efficiency of micrometastasis by sentinel node navigation surgery using indocyanine green and infrared ray laparoscopy system for gastric cancer**. *Gastric Cancer* 2012, **15**(3):287-291.
24. Lin C, Yang L: **Long Noncoding RNA in Cancer: Wiring Signaling Circuitry**. *Trends in Cell Biology* 2018, **28**(4):287-301.
25. Wen X, Han W, Liu C: **Long non-coding RNA TTTY15 silencing inhibits gastric cancer progression by sponging microRNA-98-5p to down-regulate cyclin D2 expression**. *Bioengineered* 2022, **13**(3):7380-7391.
26. Bridges MC, Daulagala AC, Kourtidis A: **LNCcation: lncRNA localization and function**. *Journal of Cell Biology* 2021, **220**(2).
27. Jia Y, Yan Q, Zheng Y, Li L, Zhang B, Chang Z, Wang Z, Tang H, Qin Y, Guan X-Y: **Long non-coding RNA NEAT1 mediated RPRD1B stability facilitates fatty acid metabolism and lymph node metastasis via c-Jun/c-Fos/SREBP1 axis in gastric cancer**. *Journal of Experimental & Clinical Cancer Research* 2022, **41**(1).
28. Yang Z, Zhi Q, Wang D, Zhang L, Preston B, Brandon C, Kuang Y, Miao R, Shi Y, Guo X: **Long Noncoding RNA C21orF96 Promotes the Migration, Invasion and Lymph Node Metastasis in Gastric Cancer**. *Anti-Cancer Agents in Medicinal Chemistry* 2016, **16**(9):1101-1108.
29. Bass AJ, Thorsson V, Shmulevich I, Reynolds SM, Miller M, Bernard B, Hinoue T, Laird PW, Curtis C, Shen H *et al*: **Comprehensive molecular characterization of gastric adenocarcinoma**. *Nature* 2014, **513**(7517):202-209.
30. Suh Y-S, Na D, Lee J-S, Chae J, Kim E, Jang G, Lee J, Min J, Ock C-Y, Kong S-H *et al*: **Comprehensive Molecular Characterization of Adenocarcinoma of the Gastroesophageal Junction Between Esophageal and Gastric Adenocarcinomas**. *Annals of Surgery* 2022, **275**(4):706-717.
31. Gan L, Lv L, Liao S: **Long non-coding RNA H19 regulates cell growth and metastasis via the miR-22-3p/Snail1 axis in gastric cancer**. *International Journal of Oncology* 2019, **54**(6):2157-2168.
32. Liu X-h, Sun M, Nie F-q, Ge Y-b, Zhang E-b, Yin D-d, Kong R, Xia R, Lu K-h, Li J-h *et al*: **Lnc RNA HOTAIR functions as a competing endogenous RNA to regulate HER2 expression by sponging miR-331-3p in gastric cancer**. *Molecular Cancer* 2014, **13**.
33. Lee C-P, Ko AM-S, Nithiyanantham S, Lai C-H, Ko Y-C: **Long noncoding RNA HAR1A regulates oral cancer progression through the alpha-kinase 1, bromodomain 7, and myosin IIA axis**. *Journal of*

Molecular Medicine-Jmm 2021, **99**(9):1323-1334.

34. Ma J, Cao K, Ling X, Zhang P, Zhu J: **LncRNA HAR1A Suppresses the Development of Non-Small Cell Lung Cancer by Inactivating the STAT3 Pathway**. *Cancers* 2022, **14**(12).
35. Fang D, Ou X, Sun K, Zhou X, Li Y, Shi P, Zhao Z, He Y, Peng J, Xu J: **m6A modification-mediated lncRNA TP53TG1 inhibits gastric cancer progression by regulating CIP2A stability**. *Cancer Science* 2022, **113**(12):4135-4150.

Figures

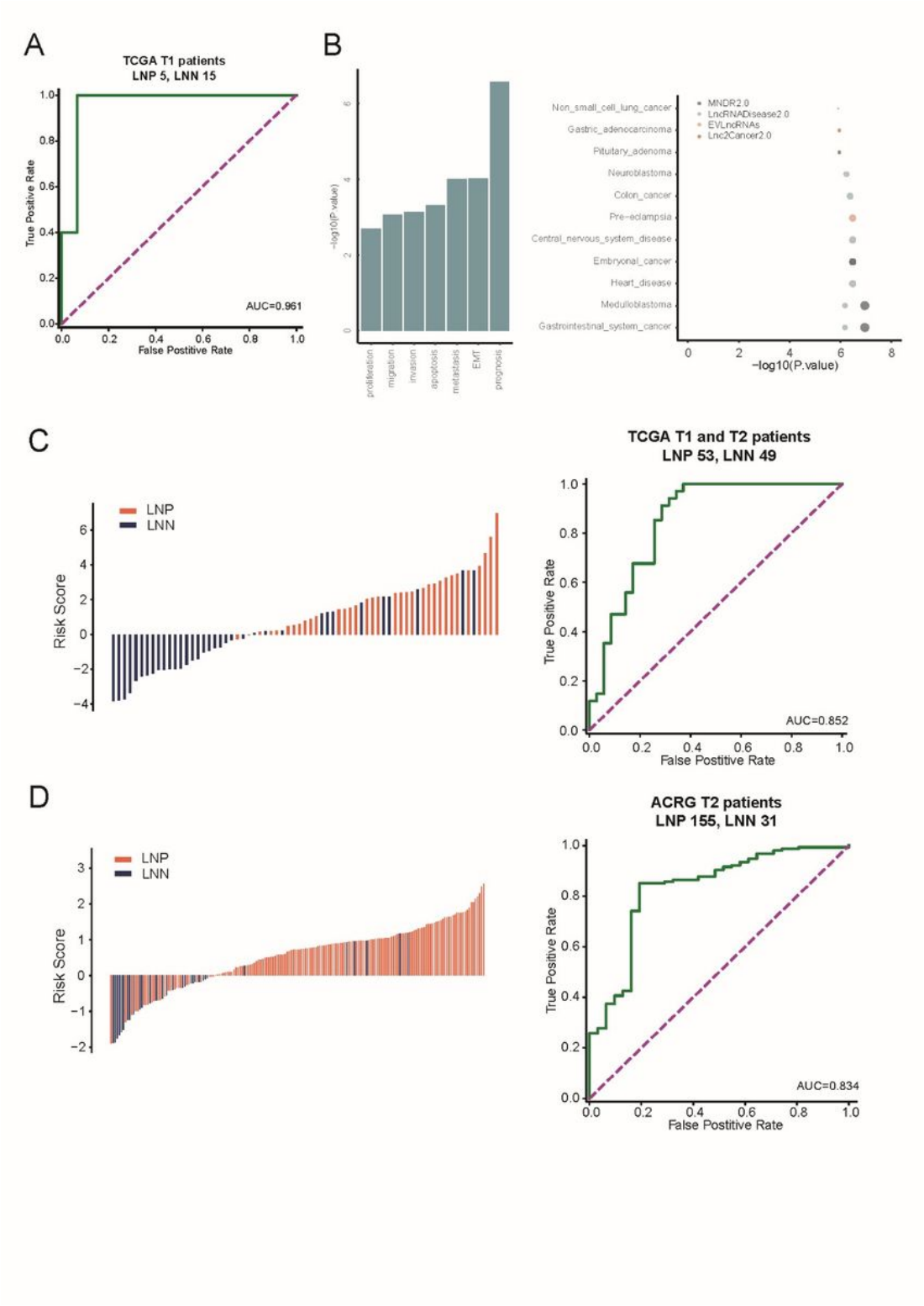

Figure 1

The long noncoding RNA (lncRNA) expression-based signature to identify lymph node metastasis in T1 and T2 stage gastric cancer

A. Receiver operating characteristic (ROC) curves revealing the diagnostic performance of the model to distinguish lymph node positive (LNP) and lymph node negative (LNN) patients in The Cancer Genome

Atlas (TCGA) T1samples. B. Histogram and bubble diagrams showing the enrichment analysis of cancer hallmark and disease in the 10-lncRNA panel. C. The lymph node (LN) risk scores divided by LN status in the T1 and T2 cohorts from the TCGA, shown as a waterfall diagram, and a receiver operating characteristic (ROC) curve showing how the 10-lncRNA risk-prediction model performed in in diagnosing patients with T1 and T2 stage disease in the TCGA data. D. The LN risk scores divided by LN status in the T2 cohort from Asian Cancer Research Group (ACRG), shown as a waterfall diagram, and a ROC curve showing how the 10-lncRNA risk-prediction model performed in in diagnosing patients T2 stage disease in the ACRG data.

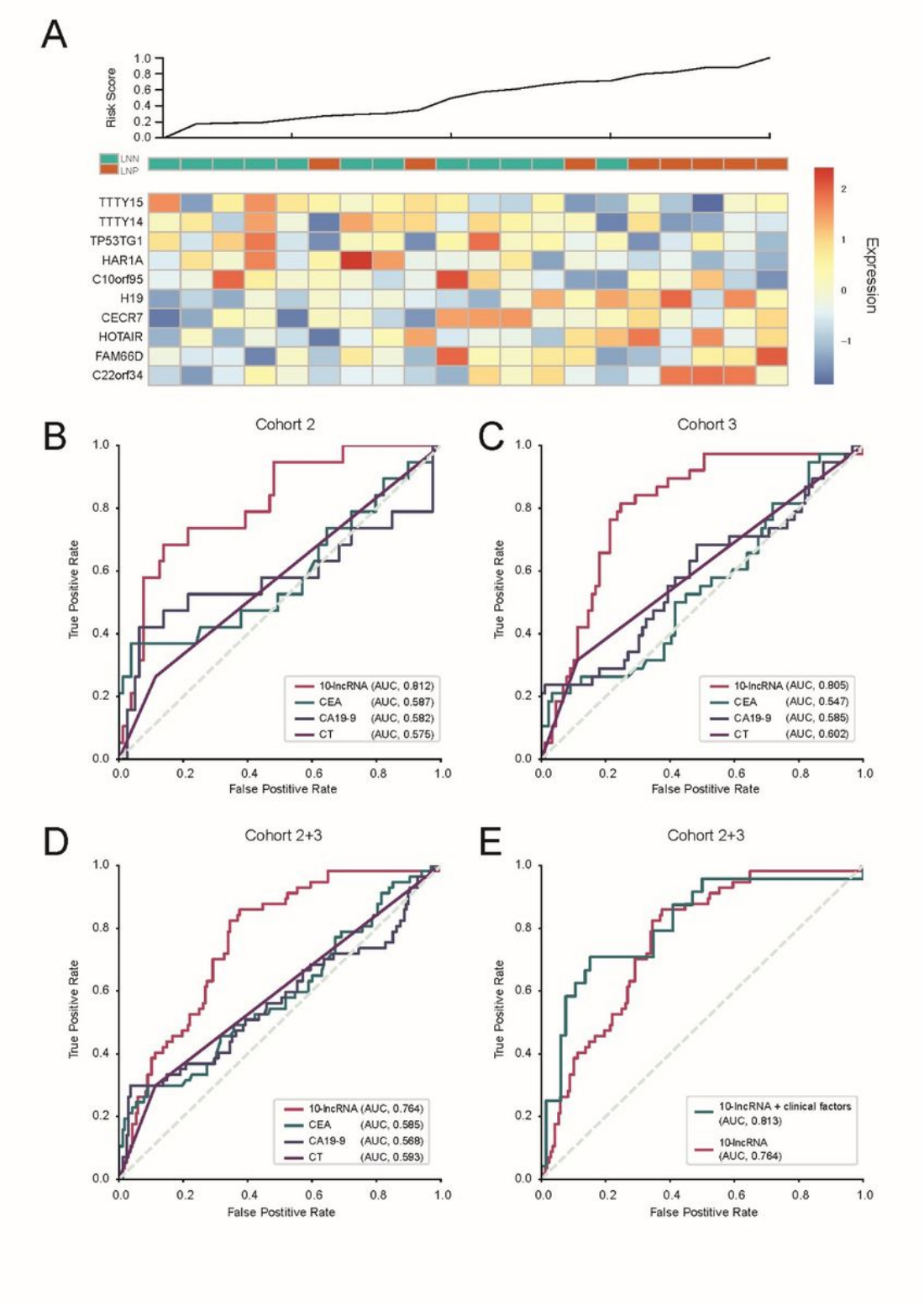

Figure 2

Performance of the 10-long noncoding RNA (lncRNA) panel to identify lymph node metastasis status in the clinical validation cohorts.

A. The risk score curve and heatmap of the lncRNAs expressed between lymph node-positive (LNP) and lymph node-negative (LNN) patients in clinical cohort 1. B and C. Receiver operating characteristic (ROC)

curves showing how the 10-lncRNA risk-prediction model performed in identifying lymph node (LN) metastasis compared with that of the carcinoembryonic antigen (CEA), carbohydrate antigen 19-9 (CA19-9), and computed tomography (CT) in clinical cohort 2 (B) ($p < 0.05$). and clinical cohort 3 (C) ($p < 0.05$), respectively. D. ROC curves illustrating the diagnostic value for identification of LN metastasis of the10-lncRNA panel compared with that of CEA, CA19-9, and CT in clinical cohort 2 and cohort 3 ($p < 0.05$). E. ROC curves illustrating the diagnostic accuracy of the combinatorial model integrating the 10-lncRNA panel and clinicopathological features in clinical cohort 2+3 compared with that of the 10-lncRNA panel alone ($p < 0.05$).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SUPtable1.xlsx
- SupplementaryFigS1.tif