

Original Paper

Microarray Expression Profile Analysis of Long Non-Coding RNAs in Human Gastric Cardiac Adenocarcinoma

Ying Wang^{a,c} Shegan Gao^{a,c} Gang Liu^a Ruinuo Jia^a Daiming Fan^b
Xiaoshan Feng^a

^aOncology Department of the First Affiliated Hospital of Henan University of Science and Technology, Luoyang, ^bState Key Laboratory of Cancer Biology and Xijing Hospital of Digestive Diseases, Fourth Military Medical University, Xi'an, China; ^cYing Wang and Shegan Gao contributed equally

Key Words

Gastric cardiac adenocarcinoma • lncRNA • Carcinogenesis • Microarray

Abstract

Background/Aims: Long noncoding RNAs (lncRNAs) are pervasively transcribed and have been shown to regulate key biological processes that maintain normal cellular functions. Abnormal regulation of these lncRNAs can promote tumorigenesis through resulting aberrant cellular essential functions. However, the roles of lncRNAs played in the development of gastric cardiac adenocarcinoma (GCA) remain unknown. With this work we aimed to show the expression profile of lncRNAs in GCA tissues compared with paired adjacent noncancerous tissue using microarray analysis in order to interrogate potential carcinogenesis molecular mechanisms of GCA from lncRNA level. **Methods:** In this study, total RNA was isolated from 15 pairs of GCA tissue, cancerous and non-cancerous, and hybridized to arraystar lncRNA V2.0 chips containing probes representing 33,000 lncRNA genes. Quantitative real-time polymerase chain reaction (PCR) was used to validate 6 up-regulated and 6 down-regulated lncRNAs. Bioinformatic analysis including gene ontology (GO) analysis, pathway analysis and network analysis was done for further investigation. **Results:** Pathway analysis indicated that 8 pathways corresponded to downregulated transcripts and that 20 pathways corresponded to up-regulated transcripts (p-value cut-off is 0.05). GO analysis showed that the highest enriched GOs targeted by up-regulated transcripts were tissue homeostasis and the highest enriched GOs targeted by the downregulated transcripts were tissue homeostasis. **Conclusion:** Our study is the first to interrogate differentially expressed lncRNAs in human GCA tissues and indicates that lncRNAs may be used as novel candidate biomarkers for the clinical diagnosis of GCA and potential targets for further therapy.

Copyright © 2014 S. Karger AG, Basel

Daiming Fan
and Xiaoshan Feng

State Key Laboratory of Cancer Biology and Xijing Hospital of Digestive Diseases, Fourth Military Medical University, Xi'an, Shan'xi (China) E-Mail fandaim@fmmu.edu.cn and Department of Oncology, Cancer Institute, the First Affiliated Hospital of Henan University of Science and Technology, Luoyang, No. 24 Jinghua Road, Luoyang, Henan (China) E-Mail samfeng137@hotmail.com

Introduction

The incidence of adenocarcinoma in esophageal and the gastroesophageal junction has increased significantly over the past decades [1], and the rates of increase vary dramatically by gender and ethnicity. The proportion of gastric cardia adenocarcinoma (GCA) has also increased strikingly in the west countries since the 1970s (rates ~5%) although the incidence of gastric cancer as a whole has decreased and in some places of northern China where is the high-risk region of upper gastrointestinal carcinomas since the 1990s [2].

The human genome sequencing showed that the human genome is pervasively transcribed into RNA and it is estimated that up to 70% of the genome is transcribed but only up to 2% of the human genome is transcribed into protein-coding RNA [3-5]. The RNA molecules lacking protein-coding potential are known as non-coding RNAs. Long non-coding RNAs refer to non-coding RNAs which are over 200 nucleotides in length [6-8]. Accumulating evidence indicates that lncRNAs are important regulators of essential cellular functions at transcriptional, post-transcriptional and translational control levels [9-13]. Aberrant regulation of these essential functions or the reactivation of some lncRNAs that are abundant during embryogenesis can promote tumor development [14, 15]. Furthermore, lncRNAs are often expressed in a spatially and temporally specific manner; that is, the lncRNAs expression often displays a cell-, tissue- and development-specific model [16], which makes lncRNAs become the potential therapeutic targets for different diseases from development to tumorigenesis and can be used as classification markers for various types of cancers [17]. Just like protein-coding genes, lncRNAs molecules may also function as tumor oncogenes or tumor suppressors through changing chromatin structure or some protein-coding genes expression level [18]. Hence, lncRNAs can be used as attractive biomarkers for tumor diagnosis and prognosis [19]. However, the roles of lncRNAs in the development of GCA and the correlation between GCA and the expression levels of these lncRNAs remains unknown.

In this study, we interrogated the differentially expression profiles of lncRNAs and mRNAs between fifteen GCA specimens and their paired adjacent noncancerous samples using microarray analysis. Our results illustrated that the lncRNA and the mRNA expression profiles differ significantly between normal gastric cardia and GCA tissues. This finding suggests that the aberrant expression levels of lncRNAs may contribute to the development and progress of GCA and that studying the differences in lncRNA expression profiles may provide new potential methods for diagnosing and treating GCA.

Materials and Methods

Patient specimens

The procedures used in this study were approved by the Institutional Review Board of the Henan University of Science and Technology and was conformed to the Helsinki Declaration, and to local legislation. A total of fifteen human primary gastric cardiac adenocarcinoma (GCA) specimens and their paired adjacent non-cancerous tissue specimens were obtained with informed consent from patients of the first affiliated hospital, Henan University of Science and Technology. All of these patients were assigned a diagnosis of GCA based on histopathology and clinical history. The anatomical sites of the primary GCA cases were from dentate line of gastroesophageal junction to 2 cm below that line. Other available clinical information that was recorded for each specimen included age, race, tumor grade, cancer stage, tissue dimensions and date (year) of resection. The pathologist assessed the tumor content by microscopic examination of cases where the percent of tumor was estimated to be $\geq 80\%$. All the patients had not received preoperative chemoradiation after which the tissue samples were taken. All the patients had not received preoperative chemoradiation after which the tissue samples were taken.

RNA isolation

Total RNA was isolated from fifteen primary GCA specimens and their paired adjacent noncancerous tissues using TRIzol reagent (Invitrogen, CA, USA). Total RNA from each specimen was quantified using a NanoDrop ND-1000 (OD 260 nm, NanoDrop, Wilmington, DE), RNA integrity was assessed using standard denaturing agarose gel electrophoresis, and the purity was judged by the ratio of absorbance at 260 nm to 280 nm (A260/A280).

DNA microarray

The Human lncRNA microarray V2.0 (manufactured by Arraystar Co. USA) was composed of lncRNAs and protein coding mRNAs in the human genome. About 33,000 lncRNAs were collected from the authoritative data sources including NCBI RefSeq, UCSC, RNADB, lncRNAs from literatures and UCRs. Sequences were selected with proprietary strategies. Repeat sequences and ncRNAs shorter than 200 bp were deleted. Each Human lncRNA array was composed of 60,302 distinct probes (60 mers) and each transcript was represented by 1–5 probes to enhance statistical confidence. Each transcript was represented by a specific exon or splice junction probe to identify individual transcripts accurately. The microarray hybridization and bioinformatic analysis was performed by KangChen Bio-tech, Shanghai, PR China.

RNA labeling and array hybridization

In this study, double-strand cDNA (ds-cDNA) was synthesized using 5 µg of total RNA by an Invitrogen SuperScript ds-cDNA synthesis kit in the presence of 100 pmol oligo dT primers. The ds-cDNA was cleaned and labeled in accordance with the NimbleGen Gene Expression Analysis protocol (NimbleGen Systems, Inc., USA). Namely, ds-cDNA was incubated with 4 µg RNase A at 37 °C for 10 min and cleaned using phenol:chloroform:isoamyl alcohol followed by ice-cold absolute ethanol precipitation. The purified cDNA was quantified using NanoDrop ND-1000 and labeled with

Cy3 using a NimbleGen One-Color DNA labeling kit according to the manufacturer's guidelines detailed in the Gene Expression Analysis protocol (NimbleGen Systems, Inc., Madison, WI, USA). One microgram of ds-cDNA was incubated for 10 min at 98°C with 1 OD of Cy3–9 mer primer. In addition, 100 pmol of deoxynucleo-side triphosphates and 100 U of the Klenow fragment (New England Biolabs, USA) were added, and the mix was incubated at 37°C for 2 h. The reaction was stopped by adding 0.1 volume of 0.5 M EDTA, and the labeled ds-cDNA was purified by isopropanol/ethanol precipitation. Microarrays were hybridized with 4 µg of Cy3-labeled ds-cDNA at 42°C during 16–20 h in NimbleGen hybridization buffer/hybridization component A in a hybridization chamber (Hybridization System-NimbleGen Systems, Inc., Madison, WI, USA). Following hybridization, washing was performed using the NimbleGen Wash Buffer kit (NimbleGen Systems, Inc., Madison, WI, USA). After being washed in an ozone-free environment, the slides were scanned using an Axon GenePix 4000B microarray scanner. The RNA labeling and microarray hybridization was performed by KangChen Bio-tech, Shanghai, PR China.

Data analysis

Informatic data analysis were performed by KangChen Biotech (Shanghai, PR China). All of the slides were scanned at 5 lm/pixel resolution using an Axon GenePix 4000B scanner (Molecular Devices Corporation) run by GenePix Pro 6.0 software (Axon). Scanned images (JPEG format) were then imported into NimbleScan software (version 2.5) for grid alignment and expression data analysis. Expression data were normalized through quantile normalization and the Robust Multichip Average (RMA) algorithm included in the NimbleScan software. The probe level files and mRNA level files were generated following normalization. All gene level files were imported into Agilent GeneSpring GX software (version 11.5.1) and normalized by the quantile method; then, Combat software was used to adjust the normalized intensity to remove batch effects. Significantly differentially expressed lncRNAs and mRNAs were identified through Volcano Plot filtering. Hierarchical clustering was performed using Agilent GeneSpring GX software (version 11.5.1) [20].

Quantitative real-time PCR (qRT-PCR)

The total RNA was isolated using TRIzol reagent (Invitrogen, CA, USA) and then reverse transcribed using PrimeScript® RT reagent Kit with gDNA Eraser (Perfect Real Time) (TaKaRa, Dalian China) according

Table 1. Clinical parameters of 15 gastric cardiac adenocarcinoma patients

Specimen No.	Age (years)	Gender (M: male /F:female)	Histological type	Histologic differentiation	T stage	Lymphatic metastasis
No.1	61	M	locally infiltrative	Poorly	T4	Yes
No.2	56	M	ulcerative	Poorly	T4	Yes
No.3	57	M	ulcerative	Moderately-Poorly	T3b	Yes
No.4	64	M	locally infiltrative	Moderately-Poorly	T4	Yes
No.5	46	M	ulcerative	Moderately-Poorly	T3a	Yes
No.6	62	M	apophysis	Moderately	T3b	Yes
No.7	57	M	ulcerative	Moderately	T3b	Yes
No.8	72	M	locally infiltrative	Moderately	T4	Yes
No.9	46	M	ulcerative	Moderately	T4	Yes
No.10	45	F	ulcerative	Moderately	T3b	Yes
No.11	63	M	ulcerative	Moderately	T3b	Yes
No.12	62	M	locally infiltrative	Moderately	T4	No
No.13	57	M	ulcerative	Moderately	T3b	Yes
No.14	77	F	apophysis	Well	T2	No
No.15	42	M	ulcerative	Well	T3b	Yes

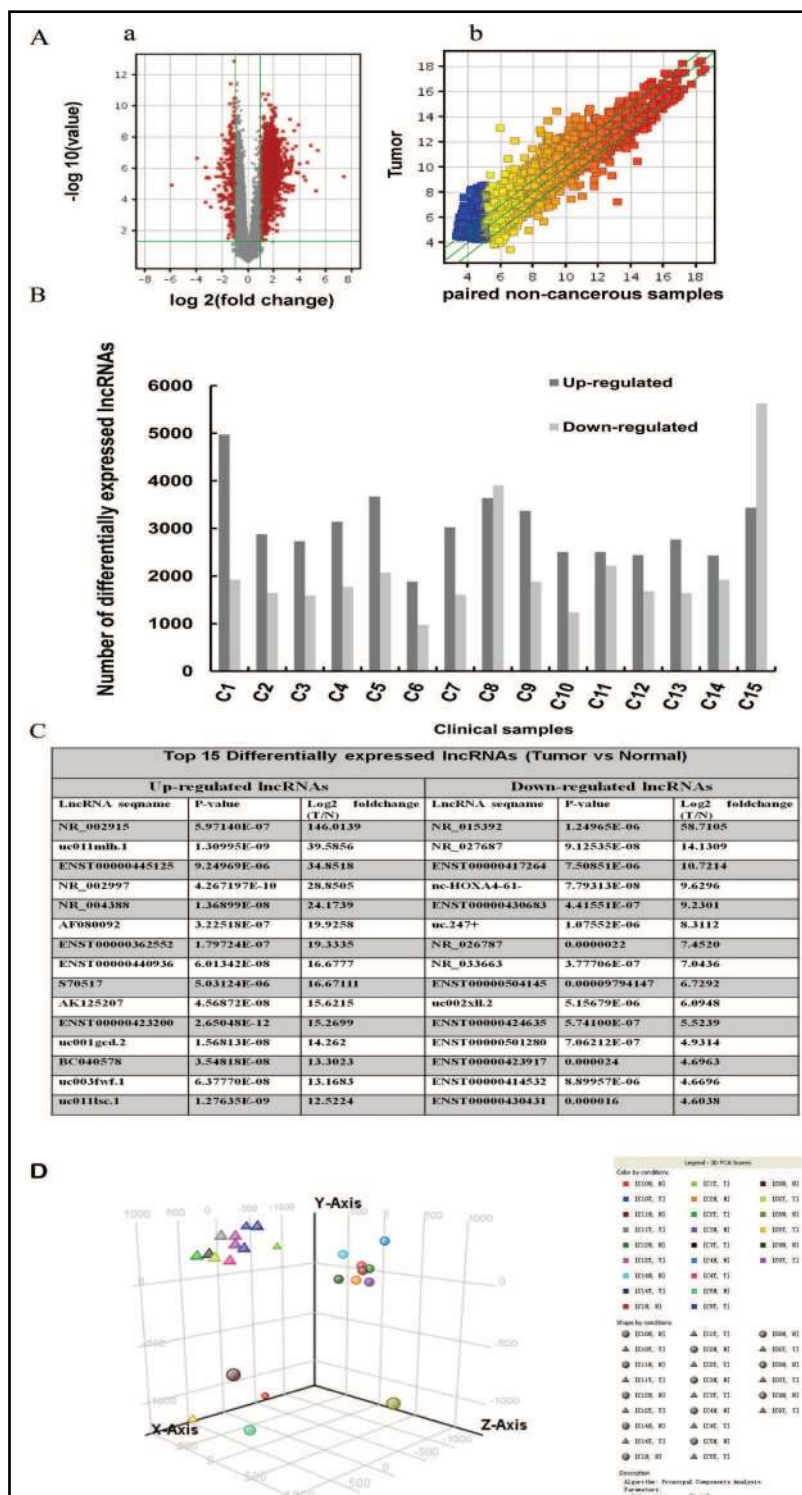
to the manufacturer's instructions. The expression of eight up-regulated lncRNAs and four down-regulated lncRNAs in fifteen patients included in this study was measured by *qRT-PCR* using SYBRGreen assays (TaKaRa, Dalian, China), and GAPDH was used as an internal control. The primers are listed in <http://cancerresearch.faisco.cn/>. The expression level of each lncRNA was represented as fold change using $2^{-\Delta\Delta Ct}$ methods. The expression level of lncRNAs differentially expressed between human primary GCA and their paired adjacent noncancerous specimens were analyzed using Student's t-test with SPSS (Version 17.0 SPSS Inc). A value of $p < 0.05$ was considered significant.

Results

Differentially expressed lncRNAs

The clinical parameters of all the patients are as shown in Table 1. The lncRNA expression profiling data showed a total of 24,842 lncRNAs expressed in gastric cardia adenocarcinoma (GCA) using microarray analysis (Fig. 1A, and <http://cancerresearch.faisco.cn/>). Based on these data, the lncRNA expression levels between fifteen human primary GCA and their paired adjacent noncancerous tissues were compared and an average of 3026 up-regulated lncRNAs (range from 1880 to 4972) and 2113 downregulated lncRNAs (range from 976 to 5629) that were significantly differentially expressed (fold change ≥ 2.0) (Fig. 1B). Among these lncRNAs, 1880 were consistently up-regulated, and 293 lncRNAs were consistently downregulated (<http://cancerresearch.faisco.cn/>). Hierarchical clustering analysis was used to arrange specimens into groups according to their expression levels, hence, we can infer the relationships among specimens. ASHG19A3A028863 (fold change:146.0139123) was the most up-regulated lncRNA, and ASHG19A3A007184 (fold change:58.71047078) was the most downregulated lncRNA (Fig. 1C). The dendrogram presents the relationships among the lncRNA expression modes between specimens (Fig. 1D). Up-regulated lncRNAs were more common than downregulated lncRNAs in our microarray data.

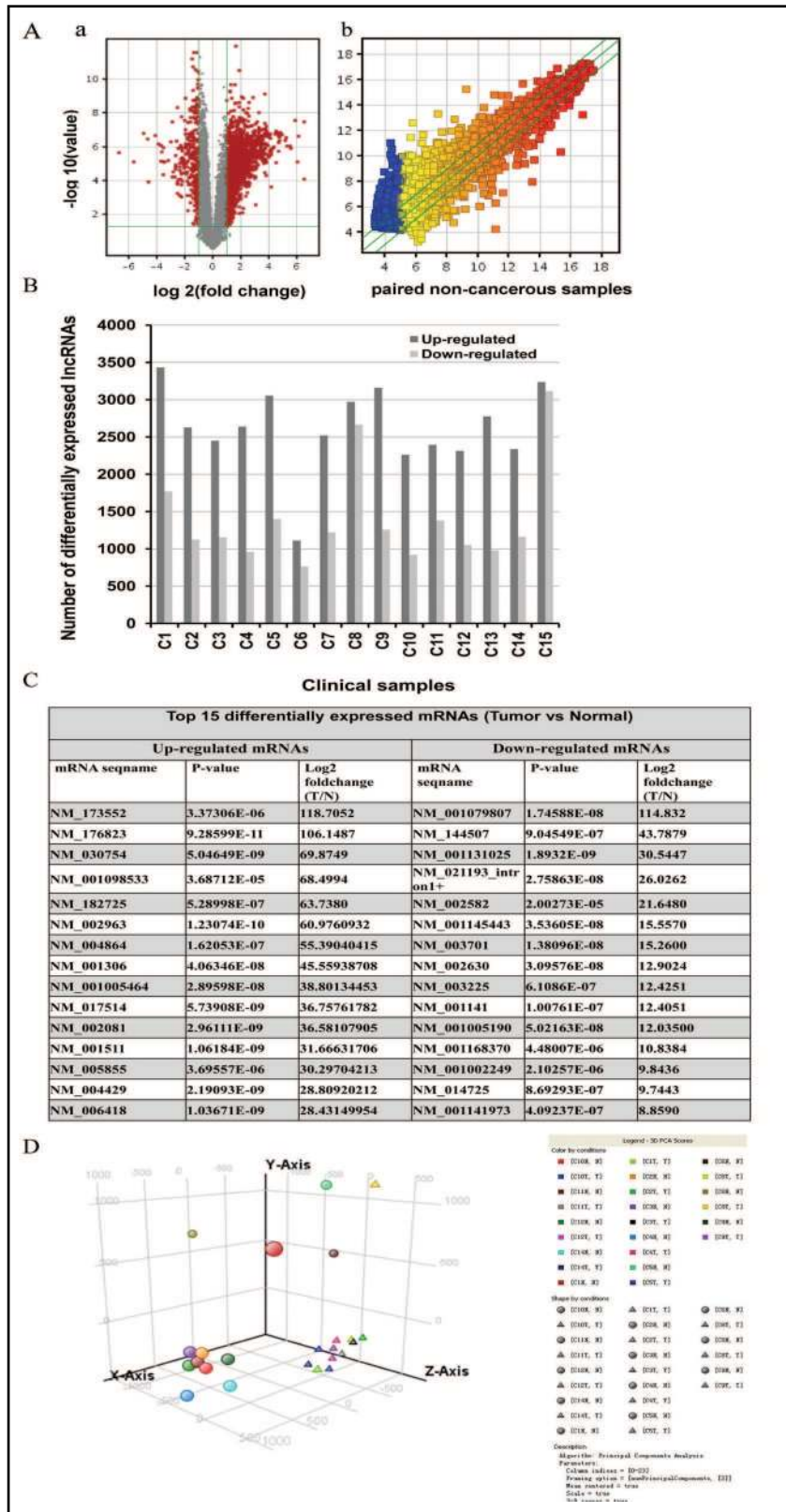
Fig. 1. A, a, Volcano Plots of lncRNA expression profile. The vertical lines correspond to 2.0-fold up and down and the horizontal line represents a P-value of 0.05. The red point in the plot represents the differentially expressed lncRNAs with statistical significance; b, The scatterplot of lncRNA expression profile, which is useful for assessing the variation (or reproducibility) between chips. B, The number of up-regulated and down-regulated lncRNAs differentially expressed between tumor and normal tissues. C, Forms showing top 10 differentially expressed lncRNAs. D, principle component analysis of lncRNA chips.



Differentially expressed mRNAs

The mRNA expression profiling data showed a total of 17,384 mRNAs identified in the GCA tissues using microarray analysis (Fig. 2A, <http://cancerresearch.faisco.cn/>). Among the fifteen human primary GCA and their paired adjacent non-cancerous samples, an average of 2619 up-regulated mRNAs (range from 1130 to 3452) and 1395 downregulated mRNAs (range from 782 to 3133) were significantly differentially expressed (fold change \geq 2.0) (Fig. 2B). In the GCA tissues, 986 mRNAs were 1735 consistently upregulated, and 257 mRNAs were

Fig. 2. A, a, Volcano Plots of mRNA expression profile; b, The scatterplot of lncRNA expression profile, which is useful for assessing the variation (or reproducibility) between chips. B, Number of up-regulated and down-regulated mRNAs differentially expressed between tumor and normal tissues in each clinical sample. C, Forms showing top 10 differentially expressed mRNAs. D, principle component analysis of mRNA chips.



consistently downregulated (<http://cancerresearch.faisco.cn/>). ASHG19A3A023341(fold change: 118.7051639) was the most up-regulated lncRNA, and ASHG19A3A046477(fold change: 114.8326712) was the most downregulated lncRNA (Fig. 2C)The Hierarchical

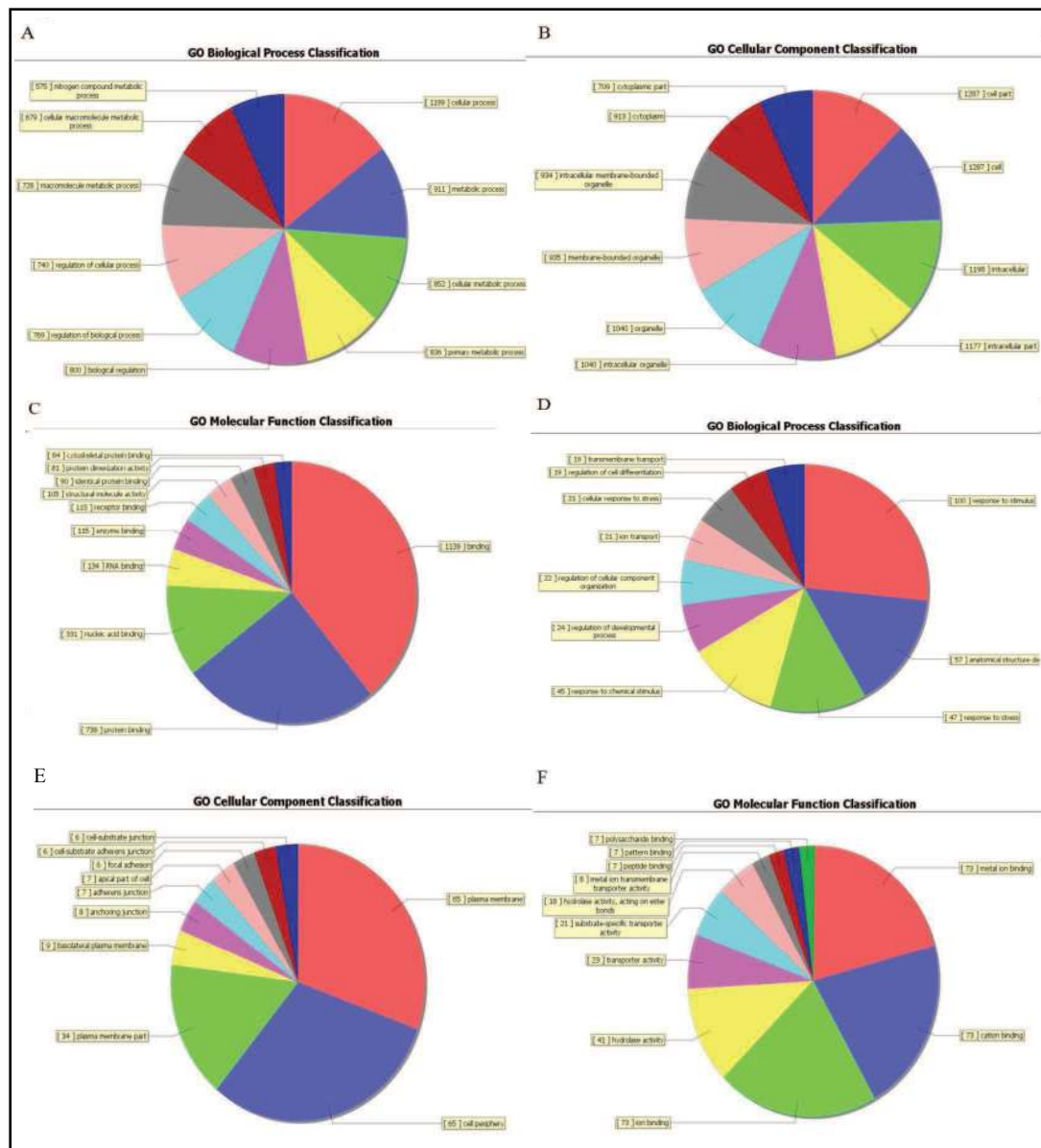


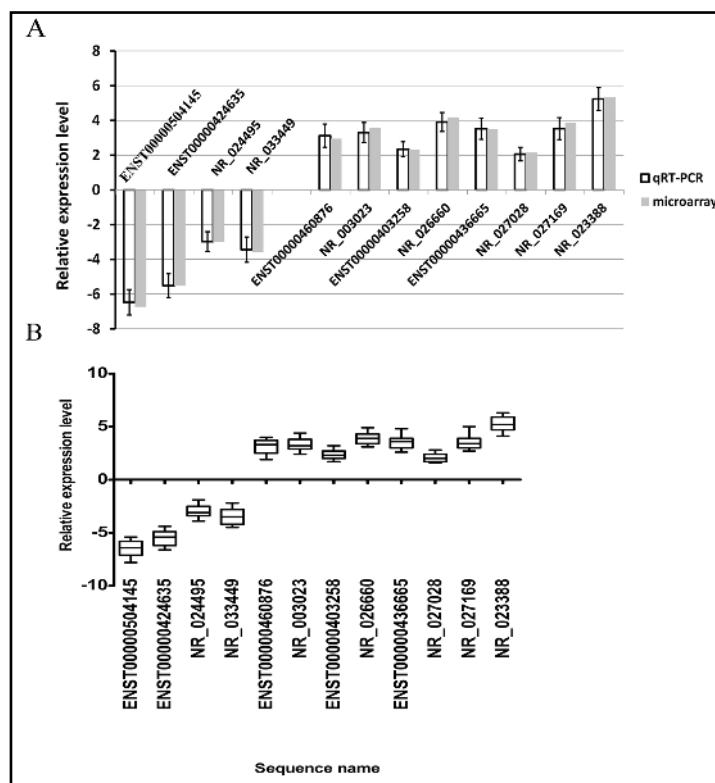
Fig. 3. The Gene Ontology project provides a controlled vocabulary to describe gene and gene product attributes in any organism (<http://www.geneontology.org>). The ontology covers three domains: Biological Process, Cellular Component and Molecular Function. Fisher's exact test is used to find if there is more overlap between the DE list and the GO annotation list than would be expected by chance. The p-value denotes the significance of GO terms enrichment in the DE genes. The lower the p-value, the more significant the GO Term (p-value<=0.05 is recommended). A-C, The highest enriched GOs targeted by up-regulated transcripts. A, biological process (BP). B, cellular component (CC). C, molecular function (MF). D-F, The highest enriched GOs targeted by down-regulated transcripts. D, biological process (BP). E, cellular component (CC). F, molecular function (MF). A-F: The chart shows the top ten counts of the significant enrichment terms.

clustering analysis presented the relationships among the mRNA expression modes that were present in the specimens (Fig. 2D).

GO analysis

In this study, Gene Ontology (GO) analysis was performed to determine the gene and gene product attributes in biological processes, cellular components and molecular functions.

Fig. 5. A, Comparison of lncRNAs expression level between microarray and qRT-PCR results. Eight up-regulated and four down-regulated differentially expressed lncRNAs were validated by qRT-PCR. The Y-axis of the columns in the chart represent the log-transformed median fold changes (T/N) in expression across twelve samples ($p < 0.05$). The qRT-PCR results are consistent with microarray data. (B) Distributions of the lncRNA expression levels ($p < 0.05$). Twelve differentially expressed lncRNAs were validated by qRT-PCR in ten human GCA and patient paired non-cancerous samples.



Term (p -value ≤ 0.05 is recommended). It has been found that the highest enriched GOs targeted by up-regulated transcripts were tissue homeostasis (GO:0001894; ontology: biological process; $p=0.0003$) (Fig. 3A), cytosolic ribosome (GO:0022626; ontology: cellular component; $p=5.13E-29$) (Fig. 3B) and structural constituent of ribosome (GO:0003735; ontology: molecular function; $p=2.80E-21$) (Fig. 3C) and that the highest enriched GOs targeted by the downregulated transcripts were tissue homeostasis (GO:0001894; ontology: biological process) (Fig. 3D), anaphase-promoting complex (GO:0005680; ontology: cellular component; $p=0.0018$) (Fig. 3E) and metal ion binding (GO:0046872; ontology: molecular function; $p=0.00057$) (Fig. 3F) (<http://cancerresearch.faisco.cn/>).

Pathway analysis

Pathway analysis indicated that 20 pathways corresponded to up-regulated transcripts (Fig. 4A) and that the most enriched network was "Ribosome - Homo sapiens (human)" (Fisher-Pvalue = $4.86E-25$) composed of 49 targeted genes. Moreover, the pathway analysis also showed that 8 pathways corresponded to downregulated transcripts (Fig. 4B) and that the most enriched network was "Collecting duct acid secretion - Homo sapiens (human)" (Fisher-Pvalue = 0.0055) composed of 3 targeted genes (<http://cancerresearch.faisco.cn/>, the recommend p -value cut-off is 0.05). Among these pathways, the gene category "Epstein-Barr virus infection" (Fig. 4C), has been reported to be involved in the carcinogenesis of gastric cardia adenocarcinoma and gastric adenocarcinoma [21-23]; the gene category "Gastric acid secretion" has been investigated as a cause of adenocarcinoma at the gastro-oesophageal junction and the distal esophagus [24, 25] and the gene category "Mismatch repair" has been showed to be involved in the development of GCA [26].

Real time quantitative PCR validation

We randomly selected eight up-regulated lncRNAs and four under-regulated lncRNAs from the differentially expressed lncRNAs. To identify the expression levels of the altered lncRNAs in GCA patients, another 12 GCA tissues and paired non-cancerous tissues were examined for the expression level of these lncRNAs using quantitative real-time PCR (qRT-

PCR). The results demonstrated that the qRT-PCR results and microarray data are consistent ($p < 0.05$, Fig. 5A and B).

Discussion

In the past decade, it has been found that the genome are widely transcribed as well as about 1% of the genome are protein-coding genes [4]. Although this pervasive transcription products named “non-coding” RNAs were once seen as transcriptional “noise,” more and more research validated that these non-coding RNAs, especially long non-coding RNAs (lncRNA) are biologically functional [7, 27]. Firstly, the abundance of lncRNAs in cancer suggests their potential role in tumorigenesis [15]; secondly, just like protein-coding genes, lncRNAs may function as tumor oncogenes or tumor suppressors through regulating related protein-coding genes expression or control transcriptional alteration directly [9]. Hence, lncRNAs could potentially be used for cancer diagnosis and prognosis [7].

Our previous research has already found that miR-645 can promote human gastric cardia adenocarcinoma (GCA) cell proliferation and inhibit GCA cell apoptosis through the down-regulation of a tumor suppressor, SOX11. However, the tumorigenesis mechanisms of GCA are far more elucidated and studies about lncRNAs dysregulation of GCA have not been reported yet.

This study showed the differential lncRNAs expression profile in GCA and non-cancerous samples using microarray analysis for the first time. We found that the lncRNA expression levels in GCA samples were altered compared with adjacent non-cancerous tissues. We analyzed fifteen human primary GCA and their paired adjacent noncancerous samples using microarray analysis and randomly selected several lncRNAs for validation by quantitative real-time PCR in 16 pairs of human primary GCA and their paired adjacent noncancerous tissue samples. The microarray expression profiles exhibited that 24,842 lncRNAs were expressed; 1880 up-regulated lncRNAs and 293 under-regulated lncRNAs were significantly differentially expressed (change fold ≥ 2.0) in fifteen GCA samples. Moreover, Gene Ontology (GO) analysis and pathway analysis were analyzed and a coexpression network was constructed to study preliminarily the biological functions and potential mechanisms of these lncRNAs in the tumorigenesis of GCA.

To our knowledge, there was not research concerning the lncRNA and mRNA microarray expression profile in human GCA to date. Most of the interest seemed to mainly focused on the single nucleotide polymorphism (SNP) of some genes in the past decade. More often, those gene SNPs might not be associated with the risk of GCA. For example, MMP-7-181 A/G [28] and MMP-3 promoter polymorphism [29] might not be associated with the development and potential of lymphatic metastasis in GCA; XRCC1 Arg194Trp and Arg399Gln SNPs might not be associated with the risk of GCA [30]; FAS-1377 G/A and FASL-844 T/C polymorphisms were not associated with the risk of GCA [31]; Lys939Gln SNP in exon 15 may have no influence on the risk of ESCC and GCA [32]; IL-10-G1082A polymorphism might not be used as a stratification marker to predicate the risk of ESCC and GCA development. However, some gene SNPs might be associated with the risk of GCA, i.e., the FAS-670 A/G genotype might decrease the risk of GCA for smoker individuals [31]; the thymidylate synthase (TS) polymorphisms might be associated with the susceptibility to GCA [33]; TNF- α -308G/A and TNF- β +252G/A genotyping may be used as a stratification markers to predicate the risk of ESCC and GCA development in North China [34] and a functional polymorphism in the matrix metalloproteinase-2 gene promoter (-1306C/T) is associated with risk of development but not metastasis of GCA [35]. We examined those genes listed above in our general microarray expression profiles of all the 15 GCA samples, and all the genes showed no significant difference in the expression between GCA tumours and adjacent non-cancerous tissues.

Because adenocarcinoma of the esophagocardia region is probably a different disease from cancer of the rest of the stomach, we then compared our findings with several other

previous research about gene expression of gastric cancer. Yu et al. have studied the gene expression profile differences in gastric cancer, pericancerous epithelium and normal gastric mucosa by gene chip [36]. We examined the genes published in that paper and found no genes were significantly differentially expressed in the GCA tissues compared with paired normal tissues; cytoplasmic expression of HER1, HER3, and HER4 were 45, 62, and 24 %, respectively in gastric cancer [37], however, the EGFR family members expression all showed no difference between GCA tissues and paired normal tissues.

It has been reported that there are epidemiological differences between oesophageal adenocarcinoma (OA) and GCA [38], suggesting that these two malignancies are separate entities with different risk factors. van Baal JW et al. have studied transcriptomes of Barrett's esophagus (BE), squamous epithelium, and gastric cardia epithelium, finding that BE proves to be an incompletely differentiated type of epithelium that shows similarities to both normal squamous and gastric cardia epithelia [39]. Zamboni et al. found that BAGE or GAGE expression was related significantly to a poor prognosis, whereas the expression of MAGE genes (in the absence of BAGE and GAGE expression) was related significantly to a good prognosis of esophageal squamous cell carcinoma and adenocarcinoma of the gastric cardia [40], however, BAGE, GAGE and MAGE were not up-regulated or not down-regulated in GCA tissues compared with normal tissues.

The finding that those genes which might be associated with development of gastric cancer or esophageal cancer did not show a significant difference between GCA and paired normal tissues may suggest that GCA is a unique entity different from esophageal cancer and gastric cancer, owing to its special developmental mechanisms.

Enhancer-like role is one of the important mechanisms by which lncRNAs could regulate some protein-coding genes transcription [41, 42], hence, enhancer lncRNAs were summarized according to their locations relative to the protein-coding genes in the genome and we found that there are 17 potential lncRNAs altogether. Genes involved in chromatin structure and epigenetic processes are sometimes used as the prognostic factor, for example, a four-gene classifier containing the non-coding gene H19, the histone HIST1H3F and the two small nucleolar RNAs, SNORA16A and SNORD14C was developed that assigns cases to low and high risk classes of squamous cell carcinomas of the larynx. In this present study, a up-regulated lncRNA, NR_026790, was found to be located upstream HIST1H3F, which encodes histone 3.1 that are responsible for the nucleosome structure of the chromosomal fiber in eukaryotes [43]. Another up-regulated lncRNA, AL832916, was found to be located upstream POLE4, a histone-fold protein that interacts with other histone-fold proteins to bind DNA in a sequence-independent manner. A up-regulated lncRNA, ENST00000430562, was found to be located upstream ZBED1, which functions as a transcription factor through binding to DNA elements found in the promoter regions of several genes related to cell proliferation, such as histone H1. hence, ZBED1 may have a role in regulating genes related to cell proliferation [44]. A up-regulated lncRNA, NR_027418, was found to be located upstream KPNA2, a uniform poor prognostic biomarker of multiple cancers such as esophagus squamous cancer, bladder cancer, gastric cancer, ovarian malignant germ cell tumor, non small cell lung cancer, infiltrative astrocytomas and so on [45].

To understand the functions of lncRNAs and the mechanisms by which lncRNAs exert their roles, pathway analysis was performed to study the differentially expressed lncRNAs and it was found that 20 pathways corresponded to up-regulated transcripts and 8 pathways corresponded to downregulated transcripts. Among these pathways, "Epstein-Barr virus infection" signal pathway has been found to be involved in the carcinogenesis of gastric cardia adenocarcinoma and gastric adenocarcinoma for Epstein-Barr virus (EBV) is a ubiquitous human herpesvirus that is associated with oncogenesis [46]. Moreover, "mismatch repair" signal pathway is involved in the development of gastric cardia adenocarcinoma. DNA mismatch repair (MMR) is a highly conserved biological pathway that plays a key role in maintaining genomic stability. MMR corrects DNA mismatches generated during DNA replication, thereby preventing mutations from becoming permanent in dividing cells.

The study using gene microarrays has helped us to understand the potential mechanisms of the carcinogenesis of gastric cardia adenocarcinoma preliminarily through "GO" analysis, signal pathway analysis and lncRNA classification analysis. Furthermore, this study for the first time focus on the molecular mechanisms of development of GCA at lncRNA level. In addition, our study showed that lncRNA NR_027418 is located near KPNA2 and ENST00000430562 is associated with ZBED1, a transcription factor involved in cell proliferation. It is suggested that lncRNAs may exert their functions through regulating the transcription of related protein-coding genes in GCA. Pathway analysis suggested the connection between signal pathways and lncRNAs which implies that lncRNAs may be used to modulate biological activities or be used as biomarkers for diagnosis or prognosis of GCA. However, to elucidate the potential roles of lncRNAs in GCA, further work including functional and mechanism experiments is needed.

Abbreviations

lncRNA (long non-coding RNA); GCA (Gastric cardiac adenocarcinoma).

Acknowledgements

Supported by National Natural Science Foundation of China, No. 81301763.

Disclosure Statement

All authors have nothing to disclose.

Reference

- 1 Vaughan TL, Davis S, Kristal A, Thomas DB: Obesity, alcohol, and tobacco as risk factors for cancers of the esophagus and gastric cardia: adenocarcinoma versus squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev* 1995;4:85-92.
- 2 Xin S, Lidong W, Zhiwei C, Fubao C, Zhong L, Xuemin L, Ran W, Zongmin F, Xin H, Zhiguo C: Distribution of ABO Blood Groups in High and Low Incidence Areas for Gastric Cardia Adenocarcinoma [J]. *J. Basic Clin Oncol* 2008;3:002.
- 3 Mattick JS, Makunin IV: Non-coding RNA. *Hum Mo Gen* 2006;15:R17-R29.
- 4 Eddy SR: Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* 2001;2:919-929.
- 5 Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A: Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* 2005;33:D121-D124.
- 6 Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai M-C, Hung T, Argani P, Rinn JL: Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010;464:1071-1076.
- 7 Mercer TR, Dinger ME, Mattick JS: Long non-coding RNAs: insights into functions. *Nat Rev Genet* 2009;10:155-159.
- 8 Ponting CP, Oliver PL, Reik W: Evolution and functions of long noncoding RNAs. *Cell* 2009;136:629-641.
- 9 Moran VA, Perera RJ, Khalil AM: Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res* 2012;40:6391-6400.
- 10 Augoff K, McCue B, Plow EF, Sossey-Alaoui K: miR-31 and its host gene lncRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer. *Mol Cancer* 2012;11:5.
- 11 Zhang B, Arun G, Mao YS, Lazar Z, Hung G, Bhattacharjee G, Xiao X, Booth CJ, Wu J, Zhang C: The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep* 2012;2:111-123.

- 12 Hung T, Chang HY: Long noncoding RNA in genome regulation: prospects and mechanisms. *RNA Biol* 2010;7:582-585.
- 13 Wang KC, Chang HY: Molecular mechanisms of long noncoding RNAs. *Mol Cell* 2011;43:904-914.
- 14 Wapinski O, Chang HY: Long noncoding RNAs and human disease. *Trends Cell Biol* 2011;21:354-361.
- 15 Nagano T, Fraser P: No-nonsense functions for long noncoding RNAs. *Cell* 2011;145:178-181.
- 16 Rinn JL, Chang HY: Genome regulation by long noncoding RNAs. *Annual Rev Biochem* 2012;81:145-166.
- 17 Lai M-c, Yang Z, Zhou L, Zhu Q-q, Xie H-y, Zhang F, Wu L-m, Chen L-m, Zheng S-s: Long non-coding RNA MALAT-1 overexpression predicts tumor recurrence of hepatocellular carcinoma after liver transplantation. *Med Oncol* 2012;29:1810-1816.
- 18 Braconi C, Kogure T, Valeri N, Huang N, Nuovo G, Costinean S, Negrini M, Miotto E, Croce C, Patel T: microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer. *Oncogene* 2011;30:4750-4756.
- 19 Han L, Zhang K, Shi Z, Zhang J, Zhu S, Zhang A, Jia Z, Wang G, Yu S: LncRNA profile of glioblastoma reveals the potential role of lncRNAs in contributing to glioblastoma pathogenesis. *Int J Oncol* 2012;40:2004.
- 20 Johnson SC: Hierarchical clustering schemes. *Psychometrika* 1967;32:241-254.
- 21 Fukayama M, Hayashi Y, Iwasaki Y, Chong J, Ooba T, Takizawa T, Koike M, Mizutani S, Miyaki M, Hirai K: Epstein-Barr virus-associated gastric carcinoma and Epstein-Barr virus infection of the stomach. *Lab Invest* 1994;71:73-81.
- 22 Shibata D, Tokunaga M, Uemura Y, Sato E, Tanaka S, Weiss L: Association of Epstein-Barr virus with undifferentiated gastric carcinomas with intense lymphoid infiltration. *Lymphoepithelioma-like carcinoma. Am J Pathol* 1991;139:469.
- 23 Oda K, Tamaru J, Takenouchi T, Mikata A, Nunomura M, Saitoh N, Sarashina H, Nakajima N: Association of Epstein-Barr virus with gastric carcinoma with lymphoid stroma. *Am J Pathol* 1993;143:1063.
- 24 Johansson K, Ask P, Boeryd B, Fransson S, Tibbling L: Oesophagitis, signs of reflux, and gastric acid secretion in patients with symptoms of gastro-oesophageal reflux disease. *Scand J Gastroentero* 1986;21:837-847.
- 25 Mearin F, Mourelle M, Guarner F, Salas A, Rtvros-Moreno V, Moncada S, Malagelada JR: Patients with achalasia lack nitric oxide synthase in the gastro-oesophageal junction. *Eur J Clin Invest* 1993;23:724-728.
- 26 Ratnasinghe LD, Abnet C, Qiao Y-L, Modali R, Stolzenberg-Solomon R, Dong Z-W, Dawsey SM, Mark SD, Taylor PR: Polymorphisms of XRCC1 and risk of esophageal and gastric cardia cancer. *Cancer Lett* 2004;216:157-164.
- 27 Gibb EA, Brown CJ, Lam WL: The functional role of long non-coding RNA in human carcinomas. *Mol Cancer* 2011;10:38-55.
- 28 Zhang J, Jin X, Fang S, Wang R, Li Y, Wang N, Guo W, Wang Y, Wen D, Wei L: The functional polymorphism in the matrix metalloproteinase-7 promoter increases susceptibility to esophageal squamous cell carcinoma, gastric cardiac adenocarcinoma and non-small cell lung carcinoma. *Carcinogenesis* 2005;26:1748-1753.
- 29 Zhang J, Jin X, Fang S, Li Y, Wang R, Guo W, Wang N, Wang Y, Wen D, Wei L: The functional SNP in the matrix metalloproteinase-3 promoter modifies susceptibility and lymphatic metastasis in esophageal squamous cell carcinoma but not in gastric cardiac adenocarcinoma. *Carcinogenesis* 2004;25:2519-2524.
- 30 Yan L, Yanan D, Donglan S, Na W, Rongmiao Z, Zhifeng C: Polymorphisms of XRCC1 gene and risk of gastric cardiac adenocarcinoma. *Dis Esophagus* 2009;22:396-401.
- 31 Zhou RM, Wang N, Chen ZF, Duan YN, Sun DL, Li Y: Polymorphisms in promoter region of FAS and FASL gene and risk of gastric cardiac adenocarcinoma. *J Gastroentero Hepat* 2010;25:555-561.
- 32 Zhou R, Li Y, Wang N, Zhang X, Dong X, Guo W: Correlation of XPC Ala499Val and Lys939Gln polymorphisms to risks of esophageal squamous cell carcinoma and gastric cardiac adenocarcinoma]. *Chin J Cancer* 2006;25:1113.
- 33 Zhang J, Cui Y, Kuang G, Li Y, Wang N, Wang R, Guo W, Wen D, Wei L, Yu F: Association of the thymidylate synthase polymorphisms with esophageal squamous cell carcinoma and gastric cardiac adenocarcinoma. *Carcinogenesis* 2004;25:2479-2485.
- 34 Guo W, Wang N, Li Y, Zhang J-H: Polymorphisms in tumor necrosis factor genes and susceptibility to esophageal squamous cell carcinoma and gastric cardiac adenocarcinoma in a population of high incidence region of North China. *Chinese Med J* 2005;118:1870-1878.

- 35 Miao X, Yu C, Tan W, Xiong P, Liang G, Lu W, Lin D: A functional polymorphism in the matrix metalloproteinase-2 gene promoter (-1306C/T) is associated with risk of development but not metastasis of gastric cardia adenocarcinoma. *Cancer Res* 2003;63:3987-3990.
- 36 Yu C-D, Xu S-H, Mou H-Z, Jiang Z-M, Zhu C-H, Liu X-L: Gene expression profile differences in gastric cancer, pericancerous epithelium and normal gastric mucosa by gene chip. *World J Gastroenterol* 2005;11:2390-12397.
- 37 Jácome AAA, Wohnrath DR, Neto CS, Carneseca EC, Serrano SV, Viana LS, Nunes JS, Martinez EZ, Santos JS: Prognostic value of epidermal growth factor receptors in gastric cancer: a survival analysis by Weibull model incorporating long-term survivors. *Gastric Cancer* 2014;17:76-86.
- 38 Huseman B: Cardia carcinoma considered as a distinct clinical entity. *Brit J Surg* 1989;76:136-139.
- 39 van Baal JW, Milano F, Rygiel AM, Bergman JJ, Rosmolen WD, van Deventer SJ, Wang KK, Peppelenbosch MP, Krishnadath KK: A comparative analysis by SAGE of gene expression profiles of Barrett's esophagus, normal squamous esophagus, and gastric cardia. *Gastroentero* 2005;129:1274-1281.
- 40 Zambon A, Mandruzzato S, Parenti A, Macino B, Dalerba P, Ruol A, Merigliano S, Zaninotto G, Zanovello P: *MAGE*, *BAGE*, and *GAGE* gene expression in patients with esophageal squamous cell carcinoma and adenocarcinoma of the gastric cardia. *Cancer* 2001;91:1882-1888.
- 41 Yang F, Zhang L, Huo Xs, Yuan Jh, Xu D, Yuan Sx, Zhu N, Zhou Wp, Yang Gs, Wang Yz: Long noncoding RNA high expression in hepatocellular carcinoma facilitates tumor growth through enhancer of zeste homolog 2 in humans. *Hepatology* 2011;54:1679-1689.
- 42 Ørom UA, Derrien T, Berlinger M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q: Long noncoding RNAs with enhancer-like function in human cells. *Cell* 2010;143:46-58.
- 43 Arents G, Burlingame RW, Wang B-C, Love WE, Moudrianakis EN: The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. *Proc Natl Acad Sci USA* 1991;88:10148-10152.
- 44 Sleumer MC, Wei G, Wang Y, Chang H, Xu T, Chen R, Zhang MQ: Regulatory elements of *Caenorhabditis elegans* ribosomal protein genes. *BMC Genomics* 2012;13:433.
- 45 Tseng S-F, Chang C-Y, Wu K-J, Teng S-C: Importin KPNA2 is required for proper nuclear localization and multiple functions of NBS1. *J Biol Chem* 2005;280:39594-39600.
- 46 Purtilo D: Epstein-Barr-virus-induced oncogenesis in immune-deficient individuals. *Lancet* 1980;315:300-303.