Construction and investigation of IncRNA-miRNA-mRNA ceRNA network in thymic epithelial neoplasms

CURRENT STATUS: UNDER REVIEW

Hanxiao Tang
Tongde Hospital Of Zhejiang Province
ORCiD: https://orcid.org/0000-0003-4609-7306

Shuyu Ge
Tongde Hospital Of Zhejiang Province

Pingping Sun
sppsuper@163.com Corresponding Author

DOI: 10.21203/rs.3.rs-18124/v1

SUBJECT AREAS
Cancer Biology Oncology

KEYWORDS
thymic epithelial neoplasms; competitive endogenous RNA network, overall survival, GO enrichment analysis, R (programming language), GEPIA
Abstract
Background Competing endogenous RNA (ceRNA) networks may be used to relate the functions of protein-coding mRNAs with those of the non-coding RNAs, such as microRNAs (miRNAs) and the long non-coding RNAs (IncRNAs). ceRNAs enable the post-transcriptional regulation of gene expression by competing for the shared miRNAs. However, the role and function of the IncRNA-miRNA-mRNA ceRNA network in thymic epithelial neoplasms (TEN) remains unknown.

Methods The miRNA, mRNA, and IncRNA expression profiles of 124 patients with TEN were downloaded from The Cancer Genome Atlas. We identified the differentially expressed (DE) miRNAs, mRNAs, and IncRNAs using the limma package in R software. The GDCRNATools package was used for the Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotations. Cytoscape software was used to construct the IncRNA-miRNA-mRNA ceRNA network. The Gene Expression Profiling Interactive Analysis platform was used to estimate the overall survival (OS) rates of the patients. Survival curves were analyzed using the log-rank test. Finally, the mRNAs in the ceRNA network were analyzed using the GOplot package in R.

Results A total of 1513, 188, and 579 TEN-specific mRNAs, IncRNAs, and miRNAs, respectively, were identified. The IncRNA-miRNA-mRNA ceRNA network was constructed, and included 53 mRNAs, 4 IncRNAs, and 27 miRNAs. A total of 10 DEmRNAs (DLX2, C8orf88, CD38, GATA3, MAL, FOXQ1, FOLH1, NLRP12, HJURP, and ACSM1) and 1 IncRNA (SNHG3) were found to be significantly associated with OS (P <0.05).

Conclusion In this study, we constructed a IncRNA-miRNA-mRNA ceRNA gene regulatory network for TEN, and identified potential prognostic and diagnostic biomarkers, as well as therapeutic targets, for the disease.

Background
Thymic epithelial neoplasms (TEN) are the most common primary tumors of the anterior mediastinum and are associated with invasion, recurrence, and metastasis involving the heart, the mediastinum, and the pleura [1] [2]. As the thymus has both immune and endocrine functions, TEN may result in the development of systemic symptoms and paraneoplastic syndromes, including myasthenia gravis,
hypohemoglobinemia, systemic lupus erythematosus, and myocarditis. Surgical treatment is the first line treatment for TEN; however, chemotherapy and radiotherapy are often used as a comprehensive treatment strategy. The treatment of TEN is challenging, and the complete removal of the tumor, the histological classification of thymoma, and paraneoplastic syndromes affect patient prognosis [3]. The morphological heterogeneity of TEN complicates the identification of molecular abnormalities specific to the thymic tumors, and there is a lack of immunohistochemical diagnostic techniques and targeted drugs. Therefore, the main goal of the present study was to investigate the molecular genetics of TEN and to identify promising therapeutic targets for the treatment of this disease. Competitive endogenous RNAs (ceRNAs), including mRNAs, pseudogenes, long non-coding (lncRNAs), and circular RNAs (circRNAs), can compete with a common set of microRNAs (miRNAs) through miRNA response elements to regulate their abundance [4]. The role of ceRNAs in gene regulation has attracted increasing attention, and may lead to important insights into tumor progression. For example, shortening of the 3’UTR region of the mRNA through an alternative polyadenylation has been shown to play a major role in suppressing the expression of the tumor suppressor genes by disrupting the ceRNA crosstalk [5].

lncRNAs and circRNAs usually regulate their expression levels through a competitive binding with miRNAs. Previous studies have revealed that lncRNAs play an important role in TEN. In the estrogen-mediated TEN mouse thymocytes, 36 differentially expressed (DE) lncRNAs affecting the cell proliferation, the cell cycle, and apoptosis were identified [6]. The FBXW7-associated lncRNAs regulate genes involved in the cell cycle and DNA repair, lymphocyte activation, and cell differentiation. Furthermore, they regulate genes controlling the response to interferon-beta in the radiation-induced thymic lymphoma in mice [7]. However, to the best of our knowledge, a lncRNA-miRNA-mRNA ceRNA network for TEN remains to be reported. Therefore, the role of the ceRNA network in TEN prognosis requires further investigation, particularly through the mining of clinical data.

The lncRNA-miRNA-mRNA ceRNA network plays an important role in the regulation of gene expression. The present study established a ceRNA network for TEN by analyzing raw data
downloaded from The Cancer Genome Atlas (TCGA). Furthermore, potential prognostic biomarkers were identified. The results of this study may shed light into the molecular mechanisms underlying the development of TEN, and may aid the identification of novel therapeutic and diagnostic biomarkers for the disease.

Methods
Patients and samples
RNAseq, miRNA, and clinical data from the TCGA-THYM study were downloaded from the TCGA database. Incomplete data (including sex, age at initial diagnosis, and survival time) were excluded. A total of 124 patients with TEN, including 94 patients with thymic cancer and 30 patients with cardiac, mediastinal or pleural metastases, were included in the present study. Sixty of the patients were females and the remaining 64 were males. A total of 115 patients survived, and nine succumbed. The number of Asian, black or African American, white, and patients with undisclosed ethnicity was 13, 6, 103, and 2, respectively.

Standardization and normalization of the RNA sequences
Data were parsed in the R software and duplicated samples were removed. Analysis of the mRNAs and lncRNAs in samples of the non-primary tumor and the non-solid tumor tissues was performed. Subsequently, the original RNAseq metadata and clinical data were merged. The merged data was normalized using the edgeR and the limma packages.

Analysis of the DE genes (DEGs)
The gdcDEAnalysis package in the R software was used to identify the DE lncRNAs, mRNAs, and miRNAs with cut-off values of |logFC| > 2 and P < 0.01. The DEGs were visualized using a bar plot and a volcano plot.

Functional enrichment analysis
The gdcEnrichAnalysis package in the R software was used to perform the Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. Bar and bubble plots were used to visualize the analyses.

Construction of the lncRNA-miRNA-mRNA ceRNA network
IncRNA-miRNA-mRNA ceRNA networks are based on the theory that IncRNAs can directly sponge miRNAs, thereby regulating their activities. The gdcCEAnalysis package in the R software was used to identify the miRNA-mRNA and the miRNA-IncRNA interactions. Thereafter, the Cytoscape software (version 3.7.2) was used to construct the IncRNA-miRNA-mRNA ceRNA network. Finally, the protein-coding genes in the ceRNA network were visualized using the GOplot package in the R software.

Survival analysis

Gene Expression Profiling Interactive Analysis (GEPIA; http://gepia.cancer-pku.cn/) is an interactive web server for cancer and normal gene expression profiling and interactive analyses [8]. To further clarify the relationship between the RNAs and survival rate of patients with TEN, GEPIA and the log-rank test (the Mantel-Cox test) were used for survival and statistical analyses, respectively. We identified the DE mRNAs and IncRNAs in the ceRNA network with significant associations with the OS, and survival curves were drawn. P < 0.05 was considered to indicate a statistically significant difference.

Results

Identification of the DE IncRNAs, miRNAs, and mRNAs

We explored 1513 TEN-specific mRNAs (1079 down-regulated and 434 up-regulated), 579 miRNAs (332 down-regulated and 247 up-regulated), and 188 IncRNAs (49 down-regulated and 139 up-regulated). The DEGs have been presented in the bar plot and the volcano plot in Figs. 1 and 2, respectively.

GO and KEGG analyses

GO analysis describes the biological processes of gene products using specific terms. The most significantly enriched biological processes of the DEGs in the present study included ‘muscle contraction’, ‘regulation of organ morphogenesis’, ‘striated muscle cell differentiation’, ‘muscle cell differentiation’, and ‘morphogenesis of an epithelium’. The most significantly enriched cellular component processes included ‘extracellular matrix’, ‘contractile fiber’, ‘sarcolemma’, ‘membrane raft’, and ‘external side of plasma membrane’. The most significantly enriched molecular function processes included ‘glycosaminoglycan binding’, ‘channel activity’, ‘passive transmembrane
transporter activity’, ‘collagen binding’, and ‘monocarboxylic acid binding’ (Fig. 3, 4). The most significantly enriched KEGG pathways included ‘dilated cardiomyopathy’, ‘hypertrophic cardiomyopathy’, ‘arrhythmogenic right ventricular cardiomyopathy’, ‘ECM-receptor interaction’, and ‘focal adhesion’ (Fig. 5).

Construction and analysis of the lncRNA-miRNA-mRNA ceRNA network

We constructed a ceRNA network based on the expression of miRNAs, lncRNAs, and mRNAs in patients with TEN. A total of 53 mRNA-, 27 miRNA-, and 4 lncRNA- nodes were identified. The network has been shown in Fig. 6. In addition, the GOCord circle graph shows the relationship between the genes and the GO terms (Fig. 7a), the GOCircle chart shows the 10 most significant GO terms (Fig. 7b), and GOBar Plot shows the P value of the relevant GO terms (Fig. 7c), as analyzed by the GOplot R package.

Survival analysis

We investigated the relationship between the expression of RNA and their clinical survival rates, especially the DEmRNA and the 4 lncRNA from ceRNA network. A total of 10 DEmRNAs (DLX2, C8orf88, CD38, GATA3, MAL, FOXQ1, FOLH1, NLRP12, HJURP, and ACSM1) and one lncRNA (SNHG3) were significantly associated with the prognosis of patients with TEN (P < 0.05; Fig. 8).

Discussion

TENs are the most common non-lymphomatous primary neoplasms of the anterior mediastinum, and are difficult to diagnose and treat. TENs have the potential to undergo malignant transformations, and complete resection and radiation therapy are only effective at the early Masaoka stages. Advanced, unresectable, or recurrent TEN are typically treated with palliative chemotherapy [9 23][10 24]. TEN type C is the most aggressive subtype of TEN and has a poor survival rate. Studies have shown that the majority of thymic carcinomas are diagnosed at advanced stages, and the cumulative incidence of recurrence at 5 years is 35% [11 1]. Variations in the morphological appearance of TEN and heterogeneous neoplastic epithelial cells confound the identification of the biological mechanisms involved in the development of TEN. Therefore, it is necessary to prospectively study the prognosis of patients with TEN. The importance of the non-coding transcriptome has become increasingly evident
in recent years [12]. The ceRNA hypothesis suggests that all RNA transcripts containing miRNA binding sites may competitively interact with miRNAs, thereby acting as ceRNAs [12 8]. ceRNAs, which include IncRNAs and mRNAs, mutually regulate each other by competing for their shared miRNAs. This type of ceRNA crosstalk has been widely reported in physiological and pathological processes [13]. As ceRNAs serve a regulatory role in tumorigenesis, they may also serve as potential prognostic biomarkers [14] [15] [16]. Thus, the mining of prognostic biomarkers by analyzing ceRNAs may aid the development of novel anti-TEN drugs, thereby improving the survival rate of patients.

In the present study, a total of 1513 DE mRNAs, 188 DE IncRNAs, and 579 DE miRNAs in TEN tissues were identified, based on the RNA-seq data obtained from TCGA. The GO and the KEGG analyses revealed that the DE mRNAs were significantly enriched in the cancer-related pathways and participated in a variety of biological processes. A total of 53 mRNAs (including CELSR3, KITLG, NTN1, MYLK, FGFR2, TGFBR3, and PLD1), 4 IncRNAs (HCG11, MAGI2-AS3, SNHG3, and NEAT1), and 27 miRNAs (including hsa-let-7e-5p, hsa-let-7f-5p, hsa-let-7 g-5p, hsa-let-7i-5p and hsa-miR-124-3p) were used to construct the ceRNA network. Additionally, 53 mRNAs from the ceRNA network were subjected to further analysis. The GOcircle and the GOcord graphs showed that these protein-coding genes were related to processes such as cell adhesion, positive regulation of the MAPK cascade and the cell surface. The associations between variations in gene expression levels and patient survival in several tumor types are widely used for evaluating the clinical importance of a given gene. On this basis, the survival curves of the 10 most significant DE mRNAs and the one IncRNA (SNHG3) were constructed, demonstrating their potential as biomarkers for predicting patient prognosis.

Accumulating evidence suggests that IncRNAs play important roles in cancer progression. Among the four IncRNAs in the ceRNA network, only SNHG3 was statistically significant. Indeed, research on the mitochondrial proteomes in ovarian cancer cells has shown that the IncRNA SNHG3 regulates energy metabolism [17]. Furthermore, SNHG3 has been revealed to promote the progression of non-small cell lung carcinoma [18] and gastric cancer [19]. However, studies have shown that the other three IncRNAs identified in the present study may also be associated with cancer. IncRNA Neat1 is essential for the paraspeckle formation and plays a vital role in inhibiting the transformation of oncogenic
signals, particularly p53. Research shows that p53 induces NEAT1 paraspeckle formation, reduces the replication-associated DNA damage. Neat1 deficiency enhances the development of a premalignant pancreatic intraepithelial neoplasia and cystic lesions in the KrasG12D-expressing mice [20]. In addition, NEAT1 targeting in the established human cancer cell lines induced a synthetic lethality with genotoxic chemotherapeutics and a nongenotoxic activation of p53 [21]. The expression of MAGI2-AS3 is down-regulated in liver cancer [22], bladder cancer [23], and breast cancer [24]. HCG11 promotes the proliferation and migration of gastric cancer cells by targeting miR-1276/CTNNB1 and activating the Wnt signaling pathway [25]. The aforementioned studies suggest that HCG11, MAGI2-AS3, SNHG3, and NEAT1 may serve as valuable prognostic markers in TEN. To the best of our knowledge, the present study was the first to investigate the role of these four lncRNAs in the diagnosis and prognosis of TEN through bioinformatics analysis. Therefore, these lncRNAs may provide important insights for the treatment of TEN in the future. It is important to note that the ceRNA prognostic biomarkers identified in this study were obtained through bioinformatics analysis, therefore future experimental studies or clinical trials will be required to validate these results.

Conclusion
The present study identified the TEN-specific IncRNAs, miRNAs, and mRNAs. A IncRNA-related ceRNA network was constructed, and IncRNAs were revealed to serve as potential prognostic biomarkers for TEN. This study may improve our understanding of the role of ceRNA regulation in the development of TEN, and lays the foundation for future clinical research.

Abbreviations
circRNAs: circular RNAs; ceRNA: Competing endogenous RNA; DEGs: DE genes; DE: differentially expressed; GEPIA: Gene Expression Profiling Interactive Analysis; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; IncRNAs: long non-coding; miRNAs: microRNAs; OS: overall survival; TCGA: The Cancer Genome Atlas; TEN: thymic epithelial neoplasms

Declarations

Acknowledgements
Not applicable.
Authors’ contributions
PS contributed to the design the research plan and make critical changes to important content. HT analyzed the data and drafts the manuscript. SG processed the chart. Each author was fully involved and reached agreement on the final manuscript version.

Funding
This research was supported by grants from Zhejiang Traditional Chinese Medicine Technology Plan (2019ZB022); National Traditional Chinese Medicine Technology Heritage Talents Training Program (T20194828003)

Availability of data and materials
All relevant data are within the manuscript.

Ethics approval and consent to participate
No ethics approval was required for this work. All utilized public data sets were generated by others who obtained ethical approval.

Consent for publication
Not Applicable.

Competing interests
The authors have declared that they have no competing of interests.

References
1. Weissferdt A, Fujimoto J, Kalhor N, Rodriguez J, Bassett R, Wistuba II, Moran CA: Expression of PD-1 and PD-L1 in thymic epithelial neoplasms. Modern Pathology. 2017, 30(6):1-8.
2. Roden AC, Eunhee SY, Jenkins SM, Edwards KK, Donovan JL, Lewis JE, Cassivi SD,
Marks RS, Garces YI, Aubry MC: Reproducibility of 3 histologic classifications and 3 staging systems for thymic epithelial neoplasms and its effect on prognosis. The American journal of surgical pathology. 2015, 39(4):427-441.

3. Carter BW, Benveniste MF, Madan R, Godoy MC, Groot PMd, Truong MT, Rosado-de-Christenson ML, Marom EM: IASLC/ITMIG staging system and lymph node map for thymic epithelial neoplasms. Radiographics. 2017, 37(3):758-776.

4. Smillie CL, Sirey T, Ponting CP: Complexities of post-transcriptional regulation and the modeling of ceRNA crosstalk. Critical reviews in biochemistry and molecular biology. 2018, 53(3):231-245.

5. Park HJ, Ji P, Kim S, Xia Z, Rodriguez B, Li L, Su J, Chen K, Masamha CP, Baillat D: 3′ UTR shortening represses tumor-suppressor genes in trans by disrupting ceRNA crosstalk. Nature genetics. 2018, 50(6):1-11.

6. Wei C, Guo D, Li Y, Zhang K, Liang G, Li Y, Ma Y, Liu J, Li Y: Profiling analysis of 17β-estradiol-regulated lncRNAs in mouse thymic epithelial cells. Physiological genomics. 2018, 50(8):553-562.

7. Snijders AM, Mao J-H: Co-Expression Network Analysis of Fbxw7-Associated LncRNAs Reveals Their Functions in Radiation-Induced Thymic Lymphoma. Insights in cancer research. 2016, 1:1-5.

8. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z: GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic acids research. 2017, 45(1):98-102.

9. Higuchi R, Goto T, Hirotsu Y, Nakagomi T, Yokoyama Y, Otake S, Amemiya K, Oyama T, Omata M: PD-L1 Expression and Tumor-Infiltrating Lymphocytes in Thymic Epithelial Neoplasms. Journal of clinical medicine. 2019, 8(11):1833.

10. Li M, Hou F, Zhao J, Zhang T, Li D, Wu W, Liu X, Xu L: Focal adhesion kinase is
overexpressed in thymic epithelial tumors and may serve as an independent prognostic biomarker. Oncology letters. 2018, 15(3):3001-3007.

11. Ahmad U, Yao X, Detterbeck F, Huang J, Antonicelli A, Filosso PL, Ruffini E, Travis W, Jones DR, Zhan Y: Thymic carcinoma outcomes and prognosis: results of an international analysis. The Journal of thoracic and cardiovascular surgery. 2015, 149(1):95-101.

12. Tay Y, Rinn J, Pandolfi PP: The multilayered complexity of ceRNA crosstalk and competition. Nature. 2014, 505(7483):344-352.

13. Zhang Y, Xu Y, Feng L, Li F, Sun Z, Wu T, Shi X, Li J, Li X: Comprehensive characterization of IncRNA-mRNA related ceRNA network across 12 major cancers. Oncotarget. 2016, 7(39):64148-64167.

14. Zhou D, Gao B, Yang Q, Kong Y, Wang W: Integrative Analysis of ceRNA Network Reveals Functional IncRNAs in Intrahepatic Cholangiocarcinoma. BioMed Research International. 2019, 2019:1-9.

15. Yan Y, Yu J, Liu H, Guo S, Zhang Y, Ye Y, Xu L, Ming L: Construction of a long non-coding RNA-associated ceRNA network reveals potential prognostic IncRNA biomarkers in hepatocellular carcinoma. Pathology-Research and Practice. 2018, 214(12):2031-2038.

16. Yang X-Z, Cheng T-T, He Q-J, Lei Z-Y, Chi J, Tang Z, Liao Q-X, Zhang H, Zeng L-S, Cui S-Z: LINC01133 as ceRNA inhibits gastric cancer progression by sponging miR-106a-3p to regulate APC expression and the Wnt/β-catenin pathway. Molecular cancer. 2018, 17(1):126.

17. Li N, Zhan X, Zhan X: The IncRNA SNHG3 regulates energy metabolism of ovarian cancer by an analysis of mitochondrial proteomes. Gynecologic oncology. 2018, 150(2):343-354.
18. Shi J, Li J, Yang S, Hu X, Chen J, Feng J, Shi T, He Y, Mei Z, He W: LncRNA SNHG3 is activated by E2F1 and promotes proliferation and migration of non-small-cell lung cancer cells through activating TGF-β pathway and IL-6/JAK2/STAT3 pathway. Journal of cellular physiology. 2020, 235(3):2891-2900.

19. Xuan Y, Wang Y: Long non-coding RNA SNHG3 promotes progression of gastric cancer by regulating neighboring MED18 gene methylation. Cell death & disease. 2019, 10(10):1-12.

20. Mello SS, Sinow C, Raj N, Mazur PK, Bieging-Rolett K, Broz DK, Imam JFC, Vogel H, Wood LD, Sage J: Neat1 is a p53-inducible lincRNA essential for transformation suppression. Genes & development. 2017, 31(11):1095-1108.

21. Adriaens C, Standaert L, Barra J, Latil M, Verfaillie A, Kalev P, Boeckx B, Wijnhoven PW, Radaelli E, Vermi W: p53 induces formation of NEAT1 IncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. Nature medicine. 2016, 22(8):1-11.

22. Yin Z, Ma T, Yan J, Shi N, Zhang C, Lu X, Hou B, Jian Z: LncRNA MAGI2-AS3 inhibits hepatocellular carcinoma cell proliferation and migration by targeting the miR-374b-5p/SMG1 signaling pathway. Journal of cellular physiology. 2019.

23. Wang F, Zu Y, Zhu S, Yang Y, Huang W, Xie H, Li G: Long noncoding RNA MAGI2-AS3 regulates CCDC19 expression by sponging miR-15b-5p and suppresses bladder cancer progression. Biochemical and biophysical research communications. 2018, 507:231-235.

24. Yang Y, Yang H, Xu M, Zhang H, Sun M, Mu P, Dong T, Du S, Liu K: Long non-coding RNA (lncRNA) MAGI2-AS3 inhibits breast cancer cell growth by targeting the Fas/FasL signalling pathway. Human cell. 2018, 31(3):232-241.

25. Zhang H, Huang H, Xu X, Wang H, Wang J, Yao Z, Xu X, Wu Q, Xu F: LncRNA HCG11
promotes proliferation and migration in gastric cancer via targeting miR-1276/CTNNB1 and activating Wnt signaling pathway. Cancer Cell International. 2019, 19(1):1-12.

Figures

Figure 1

DEG bar plot.
Figure 2

DEG volcano plot.
The most significantly enriched cellular component processes included ‘extracellular matrix’, ‘contractile fiber’, ‘sarcolemma’, ‘membrane raft’, and ‘external side of plasma membrane’. The most significantly enriched molecular function processes included ‘glycosaminoglycan binding’, ‘channel activity’, ‘passive transmembrane transporter activity’, ‘collagen binding’, and ‘monocarboxylic acid binding’
The most significantly enriched cellular component processes included 'extracellular matrix', 'contractile fiber', 'sarcolemma', 'membrane raft', and 'external side of plasma membrane'. The most significantly enriched molecular function processes included 'glycosaminoglycan binding', 'channel activity', 'passive transmembrane transporter activity', 'collagen binding', and 'monocarboxylic acid binding'.
Figure 5

The most significantly enriched KEGG pathways included ‘dilated cardiomyopathy’, ‘hypertrophic cardiomyopathy’, ‘arrhythmogenic right ventricular cardiomyopathy’, ‘ECM-receptor interaction’, and ‘focal adhesion’
Figure 6

Network
Figure 7

GOCord circle graph shows the relationship between the genes and the GO terms (Figure 7a), the GOCircle chart shows the 10 most significant GO terms (Figure 7b), and GOBar Plot shows the P value of the relevant GO terms (Figure 7c)
A total of 10 DEmRNAs (DLX2, C8orf88, CD38, GATA3, MAL, FOXQ1, FOLH1, NLRP12, HJURP, and ACSM1) and one lncRNA (SNHG3) were significantly associated with the prognosis of patients with TEN.