Identification of prognostic biomarkers of prostate cancer with long non-coding RNA-mediated competitive endogenous RNA network

GUOMEI YE, LIQIANG GUO, YANFEI XING, WENDONG SUN and MINGZHEN YUAN

Department of Urology, The Second Hospital of Shandong University, Jinan, Shandong 250033, P.R. China

Received May 31, 2018; Accepted February 5, 2019

DOI: 10.3892/etm.2019.7277

Correspondence to: Dr Mingzhen Yuan, Department of Urology, The Second Hospital of Shandong University, 247 Beiyuan Street, Jinan, Shandong 250033, P.R. China
E-mail: yuan_mingzhen@126.com

Key words: lncRNA, mRNA, ceRNA network, topology, prostate cancer
This study applied a multi-step computational approach to construct a functional LMCN associated with CRC, and the relevant lncRNAs were identified based on the constructed landscape map. Then functional enrichment analyses were carried out for mRNAs which were significantly associated with lncRNAs, and used the functions of the mature mRNAs to forecast the lncRNA functions.

Materials and methods

The expression profiles of lncRNA and mRNA in PCa. The gene expression profile was obtained from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). The dataset is no. GSE72220 which was deposited by Elai and Davicioni based on GPL575. It has a total of 147 data in the dataset, of which 90 samples are control group and 57 samples are the tumor group. The information of dataset was changed into expressionSet by computer, and stored to the working directory. The expressionSet was preprocessed by the software after finding the relevant working directory in the software analysis page, and a total of 14,451 genes were received.

Clinical characteristics of samples in the profiles. The dataset consisted of 90 normal tissues and 57 prostate tumor tissues. The characteristic of all the samples is RNA. The first step of sampling is dissecting a part of the tumor or normal tissue in the corresponding area of the slide with a sterile biopsy punch tool to extract RNAs. Then the biopsy was scraped from the slide for nucleic acid extraction. Finally, the extraction and purification of total RNA used the RNeasy FFPE kit (Qiagen, Inc., Valencia, CA, USA), and the amplification and marking of RNA used the Ovation WTA FFPE system (Nugen Technologies, Inc., San Carlos, CA, USA).

Clarification of the interaction relationship between lncRNA, miRNA and mRNA. The LMCN was constructed by integrating the expression profiles between IncRNA and mRNA and analyzing the interactions of miRNA-target. The first step is to download miRNA-mRNA interactions and IncRNA-miRNA pinsections from starBase v2.0. The starBase 2.0 is a network which decodes miRNA-ceRNA, miRNA-lncRNA and protein-RNA interactions from 108 sets of CLIP-Seq data generated from 37 independent studies. In the second step, the overlap on the mRNA and IncRNA between the expression profile and the interactions of miRNA-mRNA and IncRNA-miRNA were obtained. A new expression profile was obtained which had a total of 8,522 genes including 8,471 mRNAs and 51 lncRNAs. Finally, new interactions were collected from aforsaid miRNA-mRNA interactions and IncRNA-miRNA intersections which had 265,782 pairs of miRNA-mRNA interactions and 598 pairs of lncRNA-miRNA interactions.

Identification of ceRNA interactions. To evaluate the statistical significance of the shared miRNAs between each lncRNA and mRNA, the hypergeometric test was used for identifying competing IncRNA-mRNA interactions. Because the hypergeometric test assessment and enrichment of miRNAs have interactions of lncRNA and mRNA could achieve this goal. The formula of calculation is as follows (formula 1).

\[
P = 1 - \sum_{t=0}^{\infty} \left( \begin{array}{c} K \\ t \end{array} \right) \left( \begin{array}{c} N - K \\ M - t \end{array} \right),
\]

In formula 1, N is the total number of miRNAs, of which K and M are the numbers of miRNAs connected with the current lncRNA and mRNA, and x is the number of common miRNA shared by the lncRNA and mRNA. The calculated P-value is used to evaluate the enrichment of the function. It is noteworthy that the false discovery rate (FDR) is used to correct the P-value and the FDR <0.01 regarded as the threshold. After correction, the LMCN with P-value <0.01 contained 51 lncRNAs, 8,125 mRNAs and 34,586 interactions.

Screening the subnetworks of LMCN (sub-LMCN) regulated by the highly competitive IncRNA. To analyze the co-expression of the screened IncRNA-mRNA interactions, Pearson's correlation coefficients of the control and disease groups were calculated. The formula of calculation is as follows (formula 2).

\[
\rho_{XY} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y},
\]

In formula 2, \(\text{cov}(X,Y)\) is the covariance of variables X and Y, of which \(\sigma_X\) and \(\sigma_Y\) are the standard deviations for X and Y, respectively. FDR <0.01 was regarded as the threshold. The difference value of Pearson's correlation coefficient was >0.45 and considered as the significant coexpression of the ceRNA interactions.

Prediction of IncRNA function. The prediction of IncRNA function used the mature function of mRNA though functional enrichment analysis of the mRNAs which were significantly correlated with IncRNA. First of all, the Gene Ontology (GO) analysis was used for prediction. Using the P=0.01 as the threshold, the enrichment analysis of mRNAs in the sub-LMCN was based on the BP classification of GO analysis. The pathway was subsequently predicted. In this study, the pathway database was KEGG, and the Fisher test was used to identify the enrichment pathway of IncRNA. After eliminating the pathway with enrichment P-value <0.05, the other results were the pathway which may be regulated by IncRNA.

Results

Structure of LMCN. This study constructed an LMCN to assess the IncRNA-mediated ceRNA interaction landscape. A multi-step approach was used to accomplish this. IncRNA transcripts that compete with endogenous mRNAs to bind miRNAs could be identified by the present miRNA target prediction methods. To identify the interactions with miRNA-lncRNA and miRNA-mRNA of PCa, starBase v2.0 was applied to decode and collect data sets in large-scale CLIP-Seq data. Unlike the results predicted by current software, the data set obtained from starBase v2.0 provided high confidence miRNA-target interactions supported by the experiments. The new expression profiles of 8,522 genes were obtained, which contained 8,471 mRNAs and 51 lncRNAs after assessment of intersection between the obtained data set and the genes of the chip platform HuEx-1_0-st (Affymetrix
Human Exon 1.0 ST Array), as well as combining mRNA and lncRNA in the data set. In this new expression profile, there were 265,782 pairs of miRNA-mRNA interactions and 598 pairs of lncRNA-miRNA intersections. Then the LMCN model was constructed with the significantly co-expressed lncRNA-mRNA ceRNA pairs (Fig. 1).

Biological and topological properties revealed by LMCN. The hypergeometric test was used to evaluate the enrichment of miRNAs, and 34,586 pairs of interactions which included 51 lncRNAs and 8,125 mRNAs after screened by FDR correction were obtained. The statistics of nodes degree were obtained to reveal power law distribution, and the $R^2 = 0.999$ is the PCa-associated LMCN, a scale-free network (Fig. 2). Then the node degree and betweenness centrality (BC) of LMCN were analyzed, which are topological attributes. The box plots (Fig. 3A and B) show the statistics of the degree and BC value which included the maximum value, quartile value and median value. The scatter plots (Fig. 3C and D) show the distribution of lncRNAs and mRNAs. A higher degree demonstrated that the node was a hub, which was involved in more ceRNA interactions. A higher BC showed that the node was a bottleneck, which acted as bridges connecting different network modules. Compared to mRNA nodes, the lncRNA nodes show more specific topological properties because of their greater degrees and BC values.

Critical lncRNAs of PCa. The Pearson correlation coefficient of the screened lncRNA-mRNA interactions was calculated between the control and PCa groups using co-expression analysis. The difference value (D-value) of Pearson correlation coefficient was calculated for the same interaction between the control and PCa groups, and the interactions with D-value >0.45 were regarded as the significant coexpression ceRNA interactions. After mapping the network graph, it was found that there were 46 lncRNAs, 522 mRNAs and 569 interactions (Fig. 4). In addition, the results showed that the lncRNA nodes were always in the central region of the network, whereas the mRNA nodes were always in the outer layer. Therefore, in the LMCN of PCa, the lncRNAs dominated in the central regulatory function. Using the cytoscape software to analyze the ceRNA interactions, it was found that LINC00476, MALAT1, SNHG11, LINC00649, and ILF3-AS1 had more nodes and higher BC compared to other lncRNAs.
It suggests that these 5 lncRNAs are likely to be prognostic markers in PCa.

**Prediction of GO analysis and pathway.** In order to further verify the potential function of lncRNAs, the functional enrichment of mRNAs was analyzed, which were significantly associated with lncRNA and the function of lncRNA can be predicted by the mature function of mRNA. The mRNAs which are closely related to the central lncRNAs of PCa have an obvious enrichment in a function whose ID is GO: 0030522, and the involved genes are DAB2, ASXL1, CTNNB1, TRIM16, SAFB, DDRGK1, VDR, EZH2, TWIST1, CCNE1, HSPA1B, PUM1, NOTCH2, TRAF6, NEDD4, MEDI3, ITCH, PPARγ, MEDI4, CHUK, CNOT1, BRD8, and SRC. The pathway of the lncRNA enrichment was identified by the Fisher Test based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. After correction, the pathway with P-value <0.05 was the pathway for enrichment of mRNAs that was significantly associated with lncRNA, indicating they may be the pathway regulated by the lncRNAs (Table I).

**Discussion**

The early treatments of PCa include surgical treatment, androgen blockade and radiotherapy, but most patients still suffered from recurred tumors which will destroy the bone marrow and lead to death (11). Therefore, it is urgent to identify the molecular mechanisms to confirm potential...
The lncRNAs are implicated in regulation of PCa genera and to reveal essential laws of human diseases (12). IncRNAs are longer than 200 nucleotides but the functional open reading frames are lackig. IncRNAs have important clinical value of diagnosis, prognosis and treatment because of their biological properties, so that it has been extensively studied. The studies indicated that IncRNAs participate in a variety of cell biological processes such as cellular proliferation, differentiation, and DNA damage responses (13). At present, the aberrant expression of IncRNAs occurs in many human diseases. Aiello et al using RNA-ChIP and ChIRP found that in PCa patients HOTAIR and MALAT1 were relevant to both ERα/ERβ detectable at chromatin level (14). The results showed that MALAT1 silencing was relevant for PCa, and disclosing potential perspective manipulations in terms of transcription regulation of prognostic genes.

It has been identified that IncRNA overexpression is concerned with the progression and pathology of PCa (15). The PCA-specific LncRNAs affected PCA cell self-renewal, proliferation, survival, metastasis, and apoptosis by either transcriptional or post-transcriptional regulation. Therefore some of them are involved in distinct subtypes of PCA (16). To make clear the intermodulation relationship between IncRNA and mRNA, a functional LMCN was constructed. First the gene expression profile collected in GEO database was preprocessed, resulting in obtaining 14,451 genes. Then the miRNA-mRNA interactions and IncRNA-miRNA pinsections from starBase v2.0, were studied and new gene expression profile and interaction pairs though overlapping miRNA-mRNA and IncRNA-miRNA interaction were obtained. The new gene expression profile included 8,522 genes, and the new interaction pairs included 265,782 pairs of miRNA-mRNA interactions and 598 pairs of IncRNA-miRNA. Finally, in order to assess the significace of miRNAs shared between each mRNA and IncRNA, competitive IncRNA-mRNA interactions were identified using the hypergeometric test, and the prediction of IncRNA function used the mature mRNA function.

The IncRNAs are implicated in regulation of PCA generation and development, and their functions are accomplished by adjusting the expression level of protein-coding genes. In this study, the interaction between IncRNA and mRNA was analyzed using software, and most of the central points were IncRNA. The power law distributions shown as R²≈0.999 indicated the LMCN network has a high degree of confidence. Compared the number of nodes and BC values among different central IncRNAs, LINC00476, MALAT1, SNHG11, LINC00649 and ILF3-AS1 had the deeper interaction with more mRNAs, therefore these 5 IncRNAs were more likely to serve as the prognostic markers of PCA. MALAT1 is also named NEAT2, and was a highly conserved IncRNA. Studies have shown that it was overexpressed in multifarious human tumors and linked to enhanced cell migration, invasion and proliferation of cancers. The human tumors affected by MALAT1 included glioma (17), pancreatic (18), prostate (19) and lung cancer (20). Sebastian et al thought that the cancer metastasis was a complex process in which the cells deviated from the cancer primary site then traveled through the lymphatic or circulatory systems to become secondary tumors (21). They used PCA-osteoblast interaction microarrays to recognize the new factors which promote PCa metastasis to bone. In the result, they found that MALAT1 was not only one of the differentially expressed genes, but also regulated by Sost. It meant that the proliferation, migration, or invasion of PCa cells may be regulated by the change of MALAT1 expression. Unlike MALAT1, the other four IncRNAs had no verified experiments or studies in connection with PCa. However, it was discovered that the expression of ILF3-AS1 is negatively correlated with that of miR-200b/a/429 in melanoma tissues (22). ILF3-AS1 is a melanoma-upregulated IncRNA induces cell invasion, migration and proliferation through negatively regulating miR-200b/a/429. Chaudhry investigated the miRNA modulation mechanism in human cells treated by radiotherapy (23). SNHG11 was induced in TK6 and WTK1 cells, and its expression level was followed by a decline. SNHG11 has been well-characterized in cancer. Although the direct study between these four IncRNAs and PCa was lacking, their effects on various cancers were unquestionable. Therefore, the potential IncRNA biomarkers (e.g. LINC00476, MALAT1, SNHG11, LINC00649 and ILF3-AS1) have been inferred, which can be used for diagnosis, evaluation and gene-targeted therapy of PCa. Further studies need to be carried out for verifying this inference. The GO analysis suggested that the occurrence and progression of PCa were associated with intracellular receptor signaling pathway. The Fisher test indicated that IncRNAs may be involved in Endocytosis and Focal adhesion.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>P-value (FDR)</th>
<th>Number of mRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>4144 Endocytosis [PATH:hsa04144]</td>
<td>0.0178827965638394</td>
<td>18</td>
</tr>
<tr>
<td>4510 Focal adhesion [PATH:hsa04510]</td>
<td>0.020522962511304</td>
<td>16</td>
</tr>
<tr>
<td>4740 Offactory transduction [PATH:hsa04740]</td>
<td>0.020522962511304</td>
<td>2</td>
</tr>
<tr>
<td>5200 Pathways in cancer [PATH:hsa05200]</td>
<td>0.0196461610273246</td>
<td>26</td>
</tr>
<tr>
<td>5222 Small cell lung cancer [PATH:hsa05222]</td>
<td>0.0196461610273246</td>
<td>10</td>
</tr>
<tr>
<td>5169 Epstein-Barr virus infection [PATH:hsa05169]</td>
<td>0.020522962511304</td>
<td>16</td>
</tr>
</tbody>
</table>

GO, Gene Ontology; FDR, false discovery rate.
In conclusion, this study structured an LMCN landscape of 147 PCAs samples with a new method. The LMCN landscape was used to observe the specific topological properties and synergistic, competitive effects of IncRNAs, and it revealed regulatory interactions with coding mRNAs in PCAs.

Acknowledgements

Not applicable.

Funding

This study is supported by the Natural Science Foundation of Shandong (grant no. ZR201709290020).

Availability of data and materials

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

Authors' contributions

GY and LG conceived the study and drafted the manuscript. YX acquired the data. WS and MY analyzed the data and revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References
