Original Article
Characterization of long non-coding RNA-mediated competing endogenous RNA network to reveal potential long non-coding RNA biomarkers in rheumatoid arthritis patients under anti-TNF treatment

Yi Wang1*, Zhe Lu2*, Lei Xiao2, Fang-Long Song3, Da-Wei Jiang1, Tian-Chen Wang1, Yi-Fan Kang1, Zi-Min Wang2

1Department of Orthopaedics, Third Affiliated Hospital of Second Military Medical University, Shanghai 201805, China; 2Department of Joint and Bone Surgery, First Affiliated Hospital of Second Military Medical University, Shanghai 200433, China. *Equal contributors and co-first authors.

Received July 11, 2017; Accepted October 19, 2017; Epub December 15, 2017; Published December 30, 2017

Abstract: Objective: We designed to identify long non-coding RNA (lncRNA) signatures that predicted the therapy outcomes in rheumatoid arthritis (RA) via characterizing lncRNA-mediated competing endogenous RNA network (LMCN). Methods: Hypergeometric test was used to detect the competing lncRNA-mRNA interactions, following by co-expression analysis for the competing lncRNA-mRNA interactions relying on pearson correlation coefficient (PCC). The PCC absolute value of one interaction was defined as the weight value of one edge. Using the weight value threshold of 0.6, we established a highly competitive LMCN. To analyze the network organization, we conducted degree distribution of LMCN. Moreover, to predict the function of significant lncRNAs, we implemented the pathway analysis for its mRNA neighbors in the LMCN. Finally, we employed Biclique algorithm to extract competing modules from the LMCN. Results: Relying on a weight value > 0.6, a high-competing LMCN was constructed, which covered 50 lncRNAs, 824 mRNAs, and 926 competing endogenous (ceRNA) interactions. Based on degrees > 40, a total of 7 hub genes were identified, including SNHG12, C1RL-AS1, TTTY15, MALAT1, TAPT1-AS1, LINC00476, and LINC00649. Remarkably, a hub lncRNA LINC00476 was involved in the signaling-associated pathways (phosphatidylinositol signaling system, Ras signaling pathway, Rap1 signaling pathway, and TNF signaling pathway). Overall, we extracted 1 synergistic, competitive module containing 30 nodes. Significantly, one lncRNA LINC00476 in the LMCN was also the hub within the module. Conclusion: Our exploratory study indicates that the obtained lncRNA signatures provide novel information to better understand the mechanisms of action of anti-TNF treatment in RA patients.

Keywords: Rheumatoid arthritis, long non-coding RNA, lncRNA-mediated ceRNA network, module

Introduction
Rheumatoid arthritis (RA) is a chronic, autoimmune inflammatory disorder, which is characterized by synovitis proliferation, erosion of the marginal bone, inflammatory joint fluid, leucocyte extravasation, and systemic inflammatory manifestations. Worriedly, RA is highly associated with complications, disability, and society and economic cost [1, 2]. Anti-tumor necrosis factor (anti-TNF) agents are the most effective therapeutic schemes in RA, which serve to neutralize the signaling of pro-inflammatory cytokine [3, 4]. However, about 30% of RA patients treated by anti-TNF agents fail to obtain clinical improvement, exposing them to unnecessary side effects when their disease worsens [5, 6]. The risk for adverse effects and high costs in the non-response to anti-TNF patients have driven the seek for predictive signatures which are able to predict therapy outcome.

Though more and more efforts have been done to extract biomarkers (for example, genes coding TNF, TNF receptors, and Fc receptor IIIA) to predict response to anti-TNF in RA [7-9], predictive value has not been observed among independent cohorts. Fortunately, in recent years, many researchers have emphasized the importance of non-coding RNAs (ncRNAs) in several
Selection of lncRNA biomarkers for RA after anti-TNF treatment

biological processes [10]. Among the ncRNAs, the long ncRNAs (lncRNAs) is one of the largest gene families in the human genome [11]. Though the specific roles of lncRNAs are largely unknown, several studies have documented their key functions in many biological processes including post-transcriptional regulation and genomic imprinting [12, 13]. Of note, many studies have demonstrated that lncRNAs are involved in the occurrence and progression of RA. Two lncRNAs, FAM66C and Hotair, have been identified to regulate the immune functions in RA [14, 15]. In addition, Messsemaker et al. [16] have demonstrated that a novel lncRNA CST influences the transcript level of gene C5 which is shown to be associated with RA pathogenesis. Although the molecular roles of many lncRNAs related to RA are yet unravelled, in-depth functional and mechanistic characterization of lncRNAs which predict the response to anti-TNF therapies in RA appears to be an emerging area and a promising opportunity. Thus, uncovering the lncRNA signatures and their functions in RA patients treated by anti-TNF agents is challenging.

It is noteworthy that characterization of lncRNA functions are typically implemented relying on a ‘guilt by association’ strategy [17, 18]. In recent years, several studies have implicated lncRNAs function as competing endogenous RNAs (ceRNAs) via harboring miRNA response elements to regulate expression level of mRNAs [19, 20]. Systematic analyses of lncRNA-associated ceRNA network have been conducted in cancer [21, 22], and the results have revealed the functional significance of lncRNA-associated ceRNA network in diseases. Nevertheless, no studies have been reported the complexity and behavior of lncRNA-associated ceRNA network in RA patients treated with anti-TNF agents.

In the current analysis, in order to better predict the clinical response to anti-TNF therapy in RA, we designed to identify the candidate functional lncRNA biomarkers on the basis of establishing a lncRNA-mediated ceRNA network (LMCN). The constructed LMCN offers important clues for understanding the key roles of ceRNA mediated gene regulatory network and lncRNA signatures in predicting the response to anti-TNF in RA.

Materials and methods

Collecting miRNA-target interactions

StarBase v2.0 (http://starbase.sysu.edu.cn/) [23] offers high-quality miRNA-target interactions which are experimentally confirmed and are deposited in published studies. Herein, in our work, we derived the experimentally validated miRNA-miRNA interactions and lncRNA-miRNA interactions from StarBase 2.0 [23].

Data collection

The mRNA and lncRNA expression data of RA treated with anti-TNF were derived from the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/), which was accessible through E-GEOD-33377 [24]. The data were curated from the platform of A-AFFY-143-Affymetrix GeneChip Human Exon 1.0 ST Array version 1, [HuEx-1_0-st-v1]. Microarray data of E-GEOD-33377 included 42 RA patients, including 24 patients showing no response to the anti-TNF treatment, and 18 patients responding well to anti-TNF treatment.

Prior to analysis, we pre-processed the raw data. Specifically, the probe sets were aligned to the gene symbols, and a total of 14,451 genes were extracted. After that, these 14,451 genes were mapped to the lncRNA-miRNA interactions and miRNA-mRNA interactions. Overall, 598 lncRNA-miRNA interactions and 265,782 miRNA-mRNA interactions were identified.

Identifying potential ceRNA interactions

In our study, the hypergeometric test was used to select competing lncRNA-mRNA interactions, which can measure the significance of the common miRNAs between each mRNA and lncRNA. Assuming that the genome owned a number of H miRNAs, out of which, E and F were the numbers of miRNAs connected with the mRNA and lncRNA, and G was common miRNA number shared by lncRNA and mRNA. To assess the enrichment significance of the common miRNAs, we calculated the P values:
Then, false discovery rate (FDR) was applied to adjust the original $P$ values relying on Benjamini & Hochberg method [25]. A FDR < 0.01 was acted as the cut-off threshold.

**Construction of a highly competitive LMCN**

To evaluate the co-expression probability of IncRNA-mRNA pairs, we calculated the PCC on the basis of the expression of the competing IncRNA-mRNAs pairs according to the equation described below:

$$P_{m,n} = \frac{\sigma_{m,n}}{\sigma_m \sigma_n}$$

Where $\text{cov}(m, n)$ was the covariance of variables $m$ and $n$, while $\sigma_m$ and $\sigma_n$ respectively stood for the standard deviations (SD) for $m$ and $n$. The weight values in our work were determined as the PCC absolute value of one interaction and only edges having correlations higher than 0.6 were kept to establish the LMCN.

**Network illustration and topological analysis for LMCN**

In order to visualize the LMCN, we utilized Cytoscape tool (http://cytoscape.org/) [26] to illustrate the highly competitive LMCN. The topological centrality indexes (including degree, closeness and betweenness) are broadly employed to analyze the properties of network [27]. Of which, degree is the simplest index. As documented, degree is the number of links that one node connects with other nodes [28]. In the current work, the degree distribution of all

Figure 1. Construction of the highly competitive long non-coding RNA-mediated competitive endogenous RNA network (LMCN). The LMCN covered 50 IncRNAs, 824 mRNAs, and 926 ceRNA interactions. Blue nodes were mRNAs, and pink nodes denoted IncRNAs.
nodes in LMCN was analyzed and the nodes having degrees higher than 40 were regarded as hubs.

**Extracting synergistic, competing lncRNA modules**

Because LMCN can supply an overall view of all competing lncRNA-mRNA interactions which enable illumination of the regulatory features of the lncRNAs, the sub-networks demonstrate a more detailed picture of how lncRNAs synergized with the competing mRNAs. Biclique approach was derived from the website of the Computational Biology Laboratory in the Department of Computer Science, Iowa State University (http://genome.cs.iastate.edu/supertree/download/biclique/). Significantly, Biclique method was employed in a former study of ceRNA networks [19] to detect synergistic, competing modules in the LMCN. Thus, herein, we also used the Biclique algorithm to identify the synergistic competing modules from the LMCN.

**Analysis of functional annotation**

As reported, Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/pathway) is a database used to systematically analyze gene functions [29]. In our study, with the goal of revealing the underlying biological process of lncRNAs, we used DAVID tool to implement pathway analyses for genes in LMCN according to the ‘guilt by association’ strategy. Fisher’s test was utilized to detect the significant pathways which were selected using the criteria of FDR < 0.05.

**Results**

**Establishment of a highly competitive LMCN**

In order to construct the highly competitive LMCN, we firstly used the hypergeometric test to identify candidate lncRNA-mRNA competing pairs through evaluating the significance of the common miRNAs between each lncRNA-mRNA pair. Consequently, based on FDR < 0.01, 51 lncRNAs, 8125 mRNAs, and 34,586 ceRNA interactions were extracted. Then, we applied PCC to measure the co-expression probability of lncRNA-mRNA pairs. Relying on a weight value higher than 0.6, a high-competing LMCN was constructed, which covered 50 lncRNAs, 824 mRNAs, and 926 ceRNA interactions. The LMCN was displayed in Figure 1. From this figure, we found that the lncRNAs were in the central area of the network, but the mRNAs were typically in the outside layer (Figure 1). We also found that nodes interacted with each other, but the weight scores were different. The interactions with higher weight values might be more important than the others for RA treated by anti-TNF. The majority of interactions were distributed in the weight range of 0.6–0.7 (589), followed by the range between 0.7–0.8 (238), 0.8–0.9 (74), 0.9–1.0 (20), and > 1 (5). The weight distribution of the interactions in LMCN was shown in Figure 2.

**Degree properties of the LMCN**

To explore the organization of the LMCN in RA patients treated with anti-TNF agents, we investigated the degree distribution of LMCN. The results of the degree distribution of the entire network (R² = 0.9996) demonstrated power law distributions (Figure 3A), which demonstrated that the LMCN was a scale-free network. The investigation implicated that the LMCN was parallel to biological networks and was well-organized by a core set of lncRNA-mRNA competing principles into structured instead of random networks. Commonly, a higher degree suggested that the node was a hub which participated in more ceRNA interactions. In Figure 3B, we found that the degree distribution of the lncRNAs was greater than that of the mRNAs, which further demonstrated that although lncRNAs did not code for pro-
Selection of lncRNA biomarkers for RA after anti-TNF treatment

In order to investigate the modularity characteristics of the LMCN, the Biclique method was employed in the present work. Overall, we extracted 1 synergistic, competitive module containing 30 nodes, as exhibited in Figure 4. We observed that IncRNA LINC00476 competed with 6 mRNAs (NF1, FAM120A, SYNCRIP, CSRNP2, PSKH1, and PAWR) and 5 lncRNAs (LINC00176, MCM3AP, TP53TG1, SNHG12 and LINC00649) in a 30-ceRNA module (Figure 4), demonstrating its significant functions in RA treated with anti-TNF agents. These results also suggested that there were synergistic regulatory effects among the 6 IncRNAs. Moreover, IncRNA SNHG12 competed with 12 mRNAs and 4 IncRNAs (LINC00476, MALAT1, C1RLAS1, and LINC00649) in this module (Figure 4). Significantly, two IncRNAs SNHG12 and LINC00476 in the LMCN were also the hubs within the synergistic, competing IncRNA module.

Analysis of functional annotation

To further reveal the potential biological functions of IncRNAs in the response to anti-TNF agents in RA, KEGG pathway enrichment analysis of mRNAs in the LMCN was conducted. Based on FDR < 0.05, we found that the mRNAs were remarkably enriched in 9 pathways which were mainly involved in 4 functional clusters including neuroactive ligand-receptor interaction, signaling-related pathways (phosphatidylinositol signaling system, Ras signaling pathway, Rap1 signaling pathway, and TNF signaling pathway), metabolic process associated with fatty acid and choline, and proteoglycans in cancer. Remarkably, a hub IncRNA LINC00476 was involved in the signaling-associated pathways (phosphatidylinositol signaling system, Ras signaling pathway, Rap1 signaling pathway, and TNF signaling pathway). Moreover, another...
Selection of lncRNA biomarkers for RA after anti-TNF treatment

hub lncRNA SNHG12 was significant related to the pathway of phosphatidylinositol signaling system. Specific information was shown in Table 1.

Table 1. Significant pathways based on false discovery rate (FDR) < 0.05

<table>
<thead>
<tr>
<th>Pathway IDs</th>
<th>Pathway terms</th>
<th>FDR values</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa04080</td>
<td>Neuroactive ligand-receptor interaction</td>
<td>0.006506</td>
</tr>
<tr>
<td>hsa04070</td>
<td>Phosphatidylinositol signaling system</td>
<td>0.024879</td>
</tr>
<tr>
<td>hsa04014</td>
<td>Ras signaling pathway</td>
<td>0.049537</td>
</tr>
<tr>
<td>hsa04015</td>
<td>Rap1 signaling pathway</td>
<td>0.049537</td>
</tr>
<tr>
<td>hsa04668</td>
<td>TNF signaling pathway</td>
<td>0.049537</td>
</tr>
<tr>
<td>hsa01212</td>
<td>Fatty acid metabolism</td>
<td>0.049537</td>
</tr>
<tr>
<td>hsa04960</td>
<td>Aldosterone-regulated sodium reabsorption</td>
<td>0.049537</td>
</tr>
<tr>
<td>hsa05231</td>
<td>Choline metabolism in cancer</td>
<td>0.049537</td>
</tr>
<tr>
<td>hsa05205</td>
<td>Proteoglycans in cancer</td>
<td>0.049537</td>
</tr>
</tbody>
</table>

In order to predict the function of significant IncRNAs, we implemented the KEGG pathway analysis for its mRNA neighbors in the LMCN. We observed that inferred functions of these IncRNAs were involved in several signaling-related pathways (phosphatidylinositol signaling system, Ras signaling pathway, Rap1 signaling pathway, and TNF signaling pathway). More significantly, in our study, a hub IncRNA LINC00476 was involved in these signaling pathways. Based on the published studies, these signaling pathways are associated with immune response. Specially, phosphatidylinositol signaling is a key controller of TH17 cell differentiation, and these TH17 cells can produce IL17 which is a main driver of inflammation and autoimmunity [33, 34]. The phosphatidylinositol signaling system pathway is a component of the adaptive immune response and is essential for the maintenance of self-tolerance [35]. RAP proteins are small GTPases of the RAS family, which are expressed in many cell types. Significantly, Ras activation is required for p38 and ERK activation and MSC osteogenesis [36]. The activated ERK pathway has been reported to regulate the inflammatory response [37]. TNF-alpha is the prime inflammatory mediator, and the inflammatory factors have been demonstrated to be closely associated with the RA pathogenesis [38, 39].

Significantly, Thomson et al. [40] have demonstrated that the blood of anti-TNF non-responders may generally be associated with cytokine signaling involved in inflammation, and also might participate the specific metabolic activities related lipolysis and fatty acid oxidation. A former study has suggested that responders to anti-TNF treatment are characterized by the abnormal expression of inflammatory genes in synovial tissue [41]. Moreover, the aberrant expression of inflammatory genes in responders faster than in non-responders [42]. Baarsen et al. [43] have shown that TNF treatment causes dys-expression of genes in diverse immune related pathways including inflammation. Demonstrated herein, we infer that a hub IncRAN LINCA00476 might play impor...
tant roles in predicting the anti-TNF treatment outcome of RA patients, partially though regulating the inflammatory related signaling pathways.

However, we must take into consideration of several limitations in our study. Firstly, our study involved in the relatively small samples. Secondly, the dataset used in our analysis were downloaded from the public database, not generated by us. Moreover, there was no other independent data to verify our results because of the limitation of available lncRNA expression. Additionally, the results were obtained according to bioinformatics-based-scanning, yet were not verified using biological experiments. Thus, it is urgent to do experimental studies to validate our obtained findings.

Taken together, our study successfully extracted several lncRNA biomarkers (especially LINC00476) predictive of anti-TNF treatment outcome in RA patients. These lncRNA signatures, if verified prospectively, may have important implications for the identification of clinical outcomes of TNF blocking treatment in RA and can be viewed as a start to establish a prediction model for anti-TNF treatment outcome.

Acknowledgements

This work was supported by National Natural Science Foundation of China (no: 81371974) and National Natural Science Foundation of China (no: 81572211).

Our study were not involved in human participants or animals.

Disclosure of conflict of interest

None.

Address correspondence to: Yi-Fan Kang, Department of Orthopaedics, Third Affiliated Hospital of Second Military Medical University, 700 Moyuebi Road, Jiading District, Shanghai 201805, China. Tel: +86-021-81887671; Fax: +86-021-81887671; E-mail: kfy88_7760@126.com; Zi-Min Wang, Department of Joint and Bone Surgery, First Affiliated Hospital of Second Military Medical University, 168 Changhai Road, Yangpu District, Shanghai 200433, China. Tel: +86-021-31161708; Fax: +86-021-31161708; E-mail: dreamwangzi@yeah.net

References


Selection of lncRNA biomarkers for RA after anti-TNF treatment


[37] Zhang H, Mochhala SM and Bhatia M. Endogenous hydrogen sulfide regulates inflammato-


