Sub-pathway based approach to systematically track candidate sub-pathway biomarkers for heart failure

DONG-MEI HAN

Department of Cardiology, General Hospital of Daqing Oil Field, Daqing, Heilongjiang 163000, P.R. China

Received June 18, 2018; Accepted February 6, 2019

DOI: 10.3892/etm.2019.7319

Abstract. Identification of potential novel biomarkers for heart failure was undertaken using a sub-pathway based method. To realize this goal, heart failure-relevant dataset, reference pathways, and lncRNA-miRNA-mRNA interactions were firstly recruited. Secondly, the informative pathways were extracted relying on KEGG pathways and the mRNAs in the PCC-weighted lncRNA-mRNA interactions. Thirdly, lncRNA-regulated sub-pathways were dissected after construction of condition-specific lncRNA competitively regulated pathways (LCRP). To detect crucial heart failure-relevant lncRNAs, degree analysis was conducted for all nodes within the LCRP. Ultimately, the significance of candidate sub-pathways were assessed to further identify the significant sub-pathways. There were 44 lncRNAs, 165 mRNAs and 224 co-expressed interactions. After putting the 165 mRNAs into the reference pathways, 56 informative pathways were obtained which were then embedded into undirected graphs, and 44 lncRNAs were inserted into the pathway graphs to further construct the condition-specific LCRP. According to degree distribution, 4 hub lncRNAs were selected, including ERVK13-1, YLPM1, PDXDC2P, and LINC00482. Based on the LCRP information, a total of 36 sub-pathways mediated by lncRNAs participated in 40 complete pathways. Among these 40 pathways, we mainly concentrated on the top three sub-pathways, including a sub-part of MAPK signaling pathway, an important sub-part in ErbB signaling pathway, and a part of chemokine signaling pathway. In the top 3 significant sub-pathways, gene AKT3 was simultaneously regulated by ERVK13-1, YLPM1, and PDXDC2P. Sub-pathways including MAPK signaling pathway and hub lncRNAs (ERVK13-1, YLPM1, and PDXDC2P) may play an important role in heart failure.

Introduction

Heart failure is a costly, and fatal health problem in the world (1). In 2015, this disease affected approximately 40 million people in the world (2). Approximately 50% of the heart failure patients die in 5 years after diagnosis, but more than half of the advanced heart failure patients die within the first year following diagnosis (3). The cost to health-care of heart failure is up to 30 billion dollars per year in US alone. According to the statistical data, an urgent need is highlighted for identifying more accurate and available heart failure signatures with diagnostic and prognostic values.

Identification of gene signatures for complicated diseases such as heart failure and cancer is greatly important for diagnosis, prognosis, as well as developing personalized medicine and treatment strategy. In recent years, an increasing number of studies have proposed many biomarkers associated with heart failure. For example, cardiac troponin T, a biomarker of cardiomyocyte injury, has been demonstrated to be a predictor of side effects for heart failure patients (4). Moreover, Kusumoto et al (4) demonstrated that MYH7, ANXA2, IGFBP7, as well as DESM are promising candidate signatures of heart failure both in human and mouse. Frequently, these biomarkers cannot reflect the substrate relationship between genotypes and phenotypes, therefore they offer little information in the etiology of disorders, resulting in the urgent need of elucidating the potential pathogenesis of the disease as well as developing treatments. Additionally, reliable and simple measurements to detect the disorder earlier as well as to efficiently predict the prognosis still remain insufficient.

Thus, detecting novel biomarkers is of critical importance to understand and prevent heart failure. Long non-coding RNAs (lncRNAs) are a class of ncRNAs with >200 nucleotides (5). lncRNAs have been implicated to play an important role, such as cell cycle, differentiation, and apoptosis (6,7). Moreover, lncRNAs have been reported to competitively regulate biological pathways and to be closely related to heart-related diseases, for example, heart failure, and hypertrophy (8-10). Therefore, it is necessary to integratively analyze the joint effect of lncRNAs and pathway topologies. Moreover, extracting pathways competitively regulated by lncRNAs not only can provide novel insights into the potential mechanisms, but also help us to explore the functions of lncRNAs in a given disease (11). Currently, the functions of lncRNAs under a given disease condition have not been well revealed and a method for...
uncovering the lncRNA role is needed. Fortunately, key local
subregions, not complete pathways, have been proposed to be
more explainable to the molecular mechanisms of diseases
(12,13). It is noteworthy that detecting the sub-pathways instead
of entire pathways might be more relevant for interpreting the
related biological meaning and extracting the functions of
lncRNAs. No report on lncRNA-modulated sub-pathways in
heart failure patients is available.

In this study, to explore the etiology of heart failure, the
expression of lncRNA-mRNA and pathway topologies were
profiled to extract the lncRNAs competitively regulated sub-
pathways in heart failure using the sub-pathway based strategy.

Materials and methods

Gene expression data acquisition and pretreating. The gene
expression profile on heart failure with the accession number
E-GEOD-26887 (14) was recruited from Arrayexpress
database, which was deposited in the platform of A-Affy-
141-AffymetrixGeneChip Human Gene 1.0 AT Array
[HuGene-1-0-st-v1]. In the E-GEOD-26887 (14), there were 19
patients with heart failure, and 5 controls from left ventricle
cardiac biopsies. In the present work, a total of 24 subjects
were used for subsequent analysis.

Then, pre-treatment for the original microarray profile
was conducted. The specific steps contained background
correction using robust multiarray average (RMA) approach
(15), normalization by means of quartile (16), probe correction
via MAS, and summarization of probe values through
Medianpolish method. After probes were transformed to the
human gene symbols, the final gene expression matrix
(covering 18,715 genes) was established.

Proposed scheme. Sub-pathway based method was used to
detect sub-pathways competitively regulated by lncRNA in
heart failure. This method included four steps: i) collection of
candidate lncRNA-mRNA interactions; ii) using KEGG
pathways as backbone, construction of condition-specific
lncRNA competitively regulated pathways (LCRP) on the
basis of lncRNA data, mRNA expression data, and lncRNA-
mRNA interactions; iii) aligning competing lncRNAs and
the interesting genes into the condition-specific LCRP, then
locating sub-pathways in pathways and iv) evaluation of the
significance of candidate sub-pathways.

Collection of lncRNA-mRNA interactions. Firstly, lncRNA-
mRNA interactions were recruited from StarBase database
(version 2.0) (17), and mRNA-miRNA interactions were
collected from the databases of mir2Disease, mirTarBase,
TarBase, and miRecords (V4.0). On the basis of the common
miRNAs, we collected candidate lncRNA-mRNA competitively
regulated relationships. With the goal of guaranteeing data
reliability, we filtered and reserved several candidate
competing mRNAs for each lncRNA when they met the
following criteria. The first criterion was that we utilized
hypergeometric test to estimate the importance of the common
miRNAs, and that the threshold of false discovery rate (FDR)
was defined as 0.05. The second criterion was that Jaccard
Coefficient of the common miRNAs were ordered at top
20%. Using these criteria, 7,693 candidate lncRNA-mRNA
interactions covering 835 lncRNAs, and 1,749 mRNAs were
obtained.

After that, we took the intersection between 18,715 genes
of the microarray profile and 7,693 lncRNAs-mRNA relations-
ships, we got a total of 1,605 mRNAs as well as 94 lncRNAs.

Co-expression analysis for the lncRNA-mRNA interactions
using Pearson correlation coefficient (PCC). As known,
PCC is an index of correlation between two variables, which
ranges from -1 to 1 (18). Hence, in our study, PCC was used
to assess the co-expression possibility for any pair of lncRNA-
mRNA interactions on the basis of the matched lncRNA and
mRNA data. After analysis using Fisher’s r-to-Z transforma-
tion (19), those lncRNA-mRNA interactions that possess r
value reaching a significant positive threshold were reserved
(P<0.05).

Identification of informative pathways. KEGG is a database,
which can provide reference knowledge to comprehensively
understand cellular processes using pathway aligning (20).
Thus, in our study, we downloaded all the KEGG background
pathways based on KEGG database. After that, the mRNAs of
the reconstructed lncRNA-mRNA interactions were mapped
to the KEGG reference pathways, and then informative path-
ways were obtained. We used FDR to adjust the raw P-values
by means of Benjamini-Hochberg procedure (21). We selected
informative pathways using the criteria of FDR <0.01.

Establishment of condition-specific LCRP and hubs selection.
Then, R package was used to transform the identified informa-
tive pathways above into undirected graphs which kept the
structure of the original pathway (12). The lncRNAs in the
lncRNA-mRNA interactions reweighted by PCC were aligned
into the pathway graphs, which were as nodes via connecting to
the mRNAs regulated by lncRNAs. Ultimately, the condition-
specific LCRP were constructed, which contained lncRNA
nodes as well as lncRNA-mRNA competitively regulated
dges.

To our knowledge, hubs were frequently considered
important within the biological network. Therefore, in order
to extract significant lncRNAs, we concentrated on the hub
lncRNAs in the LCRP. In this study, the top 10% of lncRNAs
having the highest degree of connectivity were extracted and
determined as the hub lncRNAs.

Positioning sub-pathways competing regulated by lncRNAs.
LncRNAs were considered as signature nodes, since these
lncRNAs referred to the genes of interest and competing
regulation. Integration of these lncRNAs with the topology
features of LCRP is beneficial to effectively locate lncRNA-
mediated subregions. The lncRNAs were embedded in the
LCRP, and then ‘lenient distance’ similarity was used to
identify the lncRNA-regulated sub-pathways based on the
topology characteristics of the LCRP. In the current study, n
(the number of molecules) = 1 and s (the number of nodes in
the molecules in a pathway) = 8 were employed to extract the
candidate sub-pathways.

Assessment of significance of candidate sub-pathway.
Wallenius approximation approach was applied to assess the
significance of candidate sub-pathways based on BiasedUrn of R package (22), to further evaluate whether the candidate sub-pathways were competitively regulated by lncRNAs. During the evaluation process, the parameters below were used: the count of mRNAs of interest; the count of background mRNAs; the count of background mRNAs in a given sub-pathways; the number of interesting mRNAs enriched in this given sub-pathway; and the weight score of the sub-pathway. The weight was the strength of competing regulation by lncRNAs locating in this sub-pathway, which was computed using the equation described in a previous study (11). Next, FDR was used to correct the original P-values, and the significant sub-pathways were selected when on the threshold of FDR was set as 0.01.

Results

Co-expression analysis for lncRNA-mRNA interactions and informative pathways identification. Before identifying informative pathways, PCC was used to evaluate the co-expression possibility for lncRNA-mRNA interactions according to matched lncRNA as well as mRNA data. When the criteria were set at a P<0.05, overall 44 lncRNAs, 165 mRNAs and 224 co-expressed interactions were extracted (data not shown). In order to detect seed pathways, these 165 mRNAs were then embedded into the background pathways. Finally, we obtained a total of 56 informative pathways when the FDR was set at 0.01 (Table I). Of note, the three most significant pathways were PI3K-Akt signaling pathway (FDR=2.81E-10), pathways in cancer (FDR=3.64E-10), and glioma (FDR=1.02E-09).

Selecting lncRNA-regulated sub-pathways. Then, the 56 informative pathways were embedded in undirected graphs, significance of candidate sub-pathways based on BiasedUrn of R package (22), to further evaluate whether the candidate sub-pathways were competitively regulat
and the 44 lncRNAs were mapped to the pathway graphs as nodes. After that, a condition-specific LCRP was built (Fig. 1). To extract crucial lncRNAs of heart failure, a degree analysis was implemented for all nodes of the LCRP. According to degree distribution, a total of 4 hub lncRNAs were selected, including ERVK13-1, YLPM1, PDXDC2P, and LINC00482.

Based on the information of the LCRP, we discovered a total of 36 lncRNAs competitively regulating sub-pathways involved in 40 complete pathways when the FDR was <0.01 (Table II). In advanced investigation, we mainly concentrated on the top three sub-pathways (Fig. 2). The first most significant sub-pathway was path: 04010_1, which was a subpart of MAPK signaling pathway (Fig. 2A). Subsequently, we further analyzed this sub-pathway, and we found that a total of 13 lncRNAs competitively regulated this subpart synergistically. Within these lncRNAs, two lncRNAs (PDXDC2P and NEXN-AS1), directly regulated three genes. In addition, MYC was coordinately regulated by three lncRNAs (CRYM-AS1, MAL2, and UBXN8). Moreover, AKT3 was also simultaneously regulated by three lncRNAs (ERVK13-1, YLPM1, and PDXDC2P).

The second significant subregion was path: 04012-1, an important subpart in ErbB signaling pathway (Fig. 2B). Similar to the first significant sub-pathway, gene AKT3 was also simultaneously regulated by three lncRNAs (ERVK13-1, YLPM1, and PDXDC2P).

The third sub-pathway, path: 04062_2, was a part of chemokine signaling pathway (Fig. 2C). Similar to the first and the second significant sub-pathways, gene AKT3 was also simultaneously regulated by three lncRNAs (ERVK13-1,
YLPM1, and PDXDC2P). Notably, lncRNA UBXN8 competitively regulated three genes (KRAS, LYN, and SHC1). Thus, these results demonstrated that this approach could identify biological meaningful sub-pathways, and highlighted some critical lncRNAs in heart failure condition.

Discussion

Currently, alteration of lncRNA competitively regulated functions may result in the occurrence and progression of diseases, but, a better understanding of this regulatory mechanism may give more and better opportunities for developing new therapies. Hence we constructed an LCRP for the informative pathways, and integrated the lncRNAs with the topology features of LCRP. Normally, comparing with the entire pathways, sub-pathways may be more explainable to the molecular mechanisms underlying disorders. Consequently, detecting the sub-pathways which were regarded to be more possibly disrupted pathways from the LCRP would be the best means to discriminate the heart failure group from the control group.

Thus, an LCRP was constructed with 36 lncRNAs, 102 mRNA nodes and 40 pathways by combining gene expression profile, lncRNA data, pathway information and PCC related analyses. Finally, based on the information of the LCRP, a total of 36 lncRNAs competitively regulating sub-pathways involved in 40 complete pathways were identified. Among these 40 pathways, we mainly concentrated on the top three sub-pathways that competitively regulated by lncRNAs, including path: 04010_1, which was a subpart of MAPK signaling pathway, path: 04012_1, an important subpart in ErbB signaling pathway, and path: 04062_2, a part of chemokine signaling pathway.

Taking MAPK signaling pathways as examples to discuss, the MAPK signaling pathway has been implicated to control myocyte growth that responds to pathological stimuli in the heart (23). Heart failure is considered to be a remodeling process of myocyte cells, and cellular as well as extracellular
factors participate in this remodeling process to further cause the changes in myocardial structures and functions. Matrix metalloproteinases (MMPs) are the factors participating in this remodeling process. Significantly, MMPs secretion has been reported to be dependent on MAPK or MAPK signaling pathway (24,25). Thus, we speculate that a close relationship exists between the heart failure and MAPK signaling pathway.

Of note, in the first, second, and third most significant sub-pathways, gene AKT3 was simultaneously regulated by three lncRNAs (ERVK13-1, YLPM1, and PDXDC2P). Moreover, ERVK13-1, YLPM1, and PDXDC2P were all hubs in the LCRP. YLPM1 was also termed as ZAP3. A former study has demonstrated that the highest expression of ZAP3 seems to be in brain, heart and kidneys (26). Of note, ZAP3 has been implicated to decrease telomerase activity (27). Further, telomerase ablation can cause the occurrence of heart failure (28). Thus, we infer that lncRNA YLPM1 might play crucial roles in the development of heart failure, partially via regulating the telomerase activity. Although few reports have suggested the link between lncRNAs (ERVK13-1, or PDXDC2P) and heart failure, we found these lncRNAs regulated the common gene AKT3 in our study. AKT3 is one of the three isoforms of AKT which mediates cardiac growth, cell death, and myocardial angiogenesis in cardiac myocytes (29). Moreover, coronary vasculature, and cell death are essential for the maintenance of the contractile function, especially in heart failure (30). Accordingly, we infer that lncRNAs (ERVK13-1 and PDXDC2P) have indirect relationships with heart failure.

The current study had several limitations. The study population was small. The lack of verification of microarray results by means of quantitative real-time PCR technique or western blot based on animal or patient tissues was another important drawback. Thus, further studies in larger study samples are necessary to confirm these findings in terms of the efficiency as diagnostic or prognostic biomarkers of heart failure.

In conclusion, we applied sub-pathway-based analysis by introducing the concept of pathway topology, lncRNA data, and subregion of pathways. The importance of this method is that it can help to understand the etiology of diseases, and provide important insight into the functions of lncRNAs in heart failure. Based on the results, we showed that MAPK signaling pathway, ErbB signaling pathway, as well as chemokine signaling pathway were the significant sub-pathways for heart failure. Nevertheless, the verification using other datasets is implemented to demonstrate that these sub-pathways are useful in classifying heart failure and normal group.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

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

Authors’ contributions

DMH conceived the study, analyzed the data and drafted the manuscript. The author has read and approved the final manuscript.
Ethics approval and consent to participate
Not applicable.

Patient consent for publication
Not applicable.

Competing interests
The author declares that he has no competing interests.

References
1. McMurray JJ and Pfeffer MA: Heart failure. Lancet 365: 1877-1889, 2005.
2. Charlson FJ, Erskine HE, Ferrari AJ, Leung J, Whiteford HA, Abajobir AA, Knibbs LD, Lalloo R, Scott JG and Guo Y; GBD 2015 Disease and Injury Incidence and Prevalence Collaborators: Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: A systematic analysis for the Global Burden of Disease Study 2015. Lancet 388: 1545-1602, 2016.
3. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, et al; American Heart Association Statistics Committee and Stroke Statistics Subcommittee: Executive summary: heart disease and stroke statistics - 2013 update: a report from the American Heart Association. Circulation 127: 143-152, 2013.
4. Kasumoto A, Miyata M, Kubozeno T, Ikeda Y, Shinato T, Kuwahata S, Fujita S, Takasaki K, Yuasa T, Hamasaki S, et al: Highly sensitive cardiac troponin T in heart failure: Comparison with echocardiographic parameters and natriuretic peptides. J Cardiol 59: 202-208, 2012.
5. Di Gesualdo F, Capaccioli S and Lulli M: A pathophysiological view of the long non-coding RNA world. Oncotarget 5: 10976-10996, 2014.
6. Zhao W, Luo J and Jiao S: Comprehensive characterization of cancer subtype associated long non-coding RNAs and their clinical implications. Sci Rep 4: 6591, 2014.
7. Gomes da Silva AM and Silbiger VN: miRNAs as biomarkers of atrial fibrillation. Biomarkers 19: 631-636, 2014.
8. Liang WC, Fu WM, Wong CW, Wang Y, Wang WM, Hu GX, Zhang L, Xiao LJ, Wan DC, Zhang JF, et al: The IncRNA H19 promotes epithelial to mesenchymal transition by functioning as miRNA sponges in colorectal cancer. Oncotarget 6: 22513-22525, 2015.
9. Papait R, Kunderfranco P, Stirparo GG, Latronico MVG and Di Gesualdo F: A pathophysiologically mean survival time of telomere shortening as a biomarker for mortality. Hum Genet 91: 215-223, 1995.
10. Martelli F: MicroRNA dysregulation in diabetic ischemic heart disease. J Cardiovasc Transl Res 6: 876-883, 2013.
11. Pan P, Li W, Lin CH, Yang J, Shang C, Nueraberg ST, Jin K, Xu W, Lin CY, Lin CJ, et al: A long noncoding RNA protects the heart from pathological hypertrophy. Nature 514: 102-106, 2014.
12. Shi X, Xu Y, Zhang C, Feng L, Sun Z, Han J, Su F, Zhang Y, Li C and Li X: Subpathway-LNCE: Identify dysfunctional subpathways competitively regulated by IncRNAs through integrating IncRNA-miRNA expression profile and pathway topologies. Oncotarget 7: 69857-69870, 2016.
13. Li C, Han J, Yao Q, Zou C, Xu Y, Zhang C, Shang D, Zhou L, Zou C, Sun Z, et al: Subpathway-GM: Identification of metabolic subpathways via joint power of interesting genes and metabolites and their topologies within pathways. Nucleic Acids Res 41: e101, 2013.