

Identification of differentially expressed long non-coding RNAs associated with dilated cardiomyopathy using integrated bioinformatics approaches

Xiaohui Luo^{1,§}, Pengdan Luo^{2,§}, Yushun Zhang^{1,*}

¹Department of Structural Heart Disease, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China;

²Department of Obstetrics and Gynecology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China.

SUMMARY The aim of this study was to identify novel long non-coding RNA (lncRNA) biomarkers associated with dilated cardiomyopathy (DCM) and reveal the potential molecular mechanisms of DCM development using bioinformatics approaches. The array data of GSE5406, including 108 DCM samples and 16 non-failing control samples, were obtained from the Gene Expression Omnibus database. The differentially expressed lncRNAs were identified using limma package in R. Pearson's correlation analyses were performed between the differentially expressed lncRNAs and protein-coding genes based on their expression levels. Pathway enrichment of these lncRNAs was conducted based on the significantly co-expressed genes. From the receiver operating characteristic (ROC) curve, the area under the ROC curve (AUC) value was obtained and used for evaluating discriminatory ability. IDI2-AS1 and XIST were differentially expressed in DCM patients. A total of 510 co-expressed genes were identified. The enriched functions and pathways of the co-expressed genes mainly included NADH dehydrogenase activity, cardiac muscle contraction, and oxidative phosphorylation. The ROC curve analysis indicated that the two lncRNAs have favorable diagnostic values in DCM. The AUC values of XIST, IDI2-AS1, and the combination of XIST and IDI2-AS1 were 0.733 (95% CI: 0.646-0.809), 0.796 (95% CI: 0.715-0.863), and 0.823 (95% CI: 0.745-0.886), respectively. This study identified IDI2-AS1 and XIST lncRNAs and related pathways involved in the pathogenesis of DCM, thus providing potential diagnostic and therapeutic targets for DCM.

Keywords long Non-coding RNAs, biomarker, dilated cardiomyopathy, bioinformatics analysis

1. Introduction

Dilated cardiomyopathy (DCM) is the most common type of cardiomyopathy worldwide, characterized by the primary presence of left ventricular (LV) dilation and reduced systolic function in the absence of abnormal loading conditions or coronary artery disease (1,2). DCM may result in progressive end-stage heart failure and sudden cardiac arrest, which is closely associated with excessive morbidity and premature mortality (3). Despite the rapid advancement in therapeutic modalities, the 5-year mortality for DCM remains up to 20%, and most of them die from heart failure and ventricular arrhythmias (4). Moreover, patients with DCM are usually asymptomatic at the early stage, and the period of diagnosis and treatment may be considerably delayed (5). In the majority of cases of idiopathic DCM, the etiology remains still unknown.

Recently, an increasing number of studies have confirmed that lncRNAs play important roles in the heart, including DCM-induced chronic heart failure initiation and progression at the posttranscriptional level (6). LncRNAs represent a cluster of RNAs that are >200 nucleotides in length without protein coding potential (7). LncRNAs can perform multiple biological functions, such as RNA processing, regulation of chromatin remodeling, marker of cell fate, and as a competing endogenous RNA (8-11). Previous studies have reported that lncRNAs play critical roles in the progression as well as diagnostic and therapeutic targets of cardiovascular diseases (12,13). Nevertheless, further studies required to explore the potential functions of lncRNAs in the development of DCM. In this study, we aimed to explore more lncRNAs involved in DCM. We screened differentially expressed lncRNAs based on GSE5406, followed by a series of bioinformatics analyses to identify key lncRNAs involved in DCM.

2. Methods

2.1. Microarray gene expression

GSE5406 gene expression profile was obtained from Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) database, based on the platform of the PL96 Affymetrix Human Genome U133A expression beadchip. The dataset was an mRNA expression profile chip, which contained 16 non-failing control samples and 108 samples from DCM patients. The platform contains both mRNAs and lncRNAs; therefore, we obtained the lncRNA expression profile by repurposing the probes in the mRNA expression profiles to lncRNAs according to the annotation in the GENCODE version GRCh38 (<http://www.gencodegenes.org>). The probe without a gene symbol was removed. For different multiple probe sets mapping to the same mRNA or lncRNA, the mean value of different probes was selected as the final expression value.

2.2. Differentially expressed lncRNA identification

The LIMMA package (Linear Models for Microarray Data) in R software (version 3.5.1) with multiple testing corrections based on the Benjamini & Hochberg method was used to screen out the differentially expressed lncRNAs between DCM and control groups (14). Adjusted false discovery rate p values < 0.05 , along with fold change ≥ 1.5 , were used as the thresholds for difference analysis. Then, cluster analysis was carried out using the Pheatmap package in R to further analyze the differentially expressed lncRNAs.

2.3. lncRNA-related protein-coding gene identification and enrichment analysis

The Pearson's correlation coefficients between the differentially expressed lncRNAs and the protein-coding genes were calculated to determine co-expression relationships (15). The protein-coding genes positively or negatively correlated with the differentially expressed lncRNAs were considered as lncRNA-related protein-coding genes ($|\text{Pearson's correlation coefficient}| > 0.6$ and $p\text{-value} < 0.001$). Gene Ontology (GO) functional enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were also conducted using the biological tool. The clusterProfiler package was used to explore the biological meaning and pathways behind numerous genes (16).

2.4. Diagnostic value of differentially expressed lncRNAs in DCM

To test the predictive value of the differentially expressed lncRNAs, we generated a receiver operating

characteristic (ROC) curve using the lncRNA expression data from 108 DCM patients and 16 controls. The area under the ROC curve (AUC) value was utilized to determine the diagnostic effectiveness in discriminating patients with DCM patients from normal individuals.

3. Results

3.1. Identification of differentially expressed lncRNAs

According to the selection criteria, a total of 107 lncRNAs were obtained, and two differentially expressed lncRNAs were identified between the DCM and control groups, including upregulated IDI2-AS1 and downregulated XIST. As shown in the heat map (Figure 1), the differentially expressed lncRNAs were predominantly separated into two clusters.

3.2. Functional enrichment analysis of pathways correlated with the identified lncRNAs in DCM

We examined the expression correlation between identified lncRNAs and genes by calculating the Pearson's correlation coefficient through expression profiles in 108 patients with DCM. Co-expression between 510 genes and the two lncRNAs (Pearson's correlation coefficient > 0.6 and $p\text{-value} < 0.001$) was found. The highest Pearson's correlation coefficient of determination between two lncRNAs and mRNA matrix is shown in Figure 2. We performed GO and KEGG enrichment analyses to uncover specific functional categories of the co-expressed genes. As a result, 510 co-expressed genes clustered most significantly in NADH dehydrogenase activity, cytochrome-c oxidase activity, cofactor binding, and oxidoreductase activity (Figure 3A). KEGG pathway enrichment showed that genes were mainly involved in oxidative phosphorylation and cardiac muscle contraction (Figure 3B).

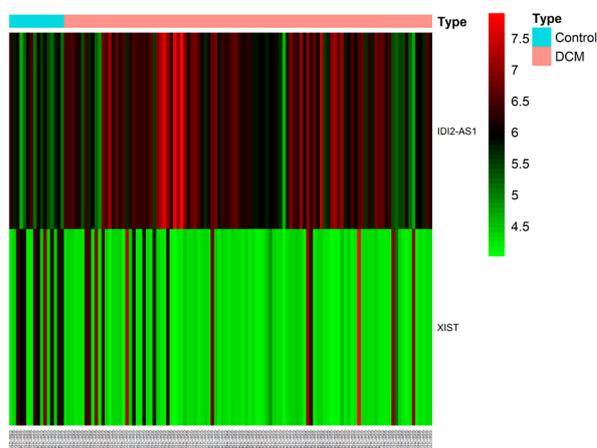


Figure 1. Heatmap results of differentially expressed long non-coding RNAs.

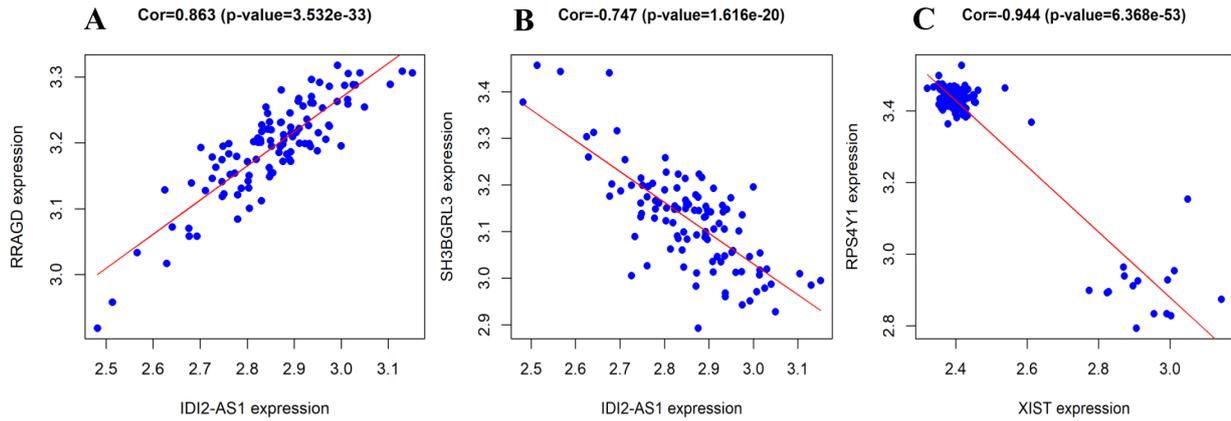


Figure 2. The highest Pearson's correlation coefficient of determination between two long non-coding RNAs and the mRNA matrix.

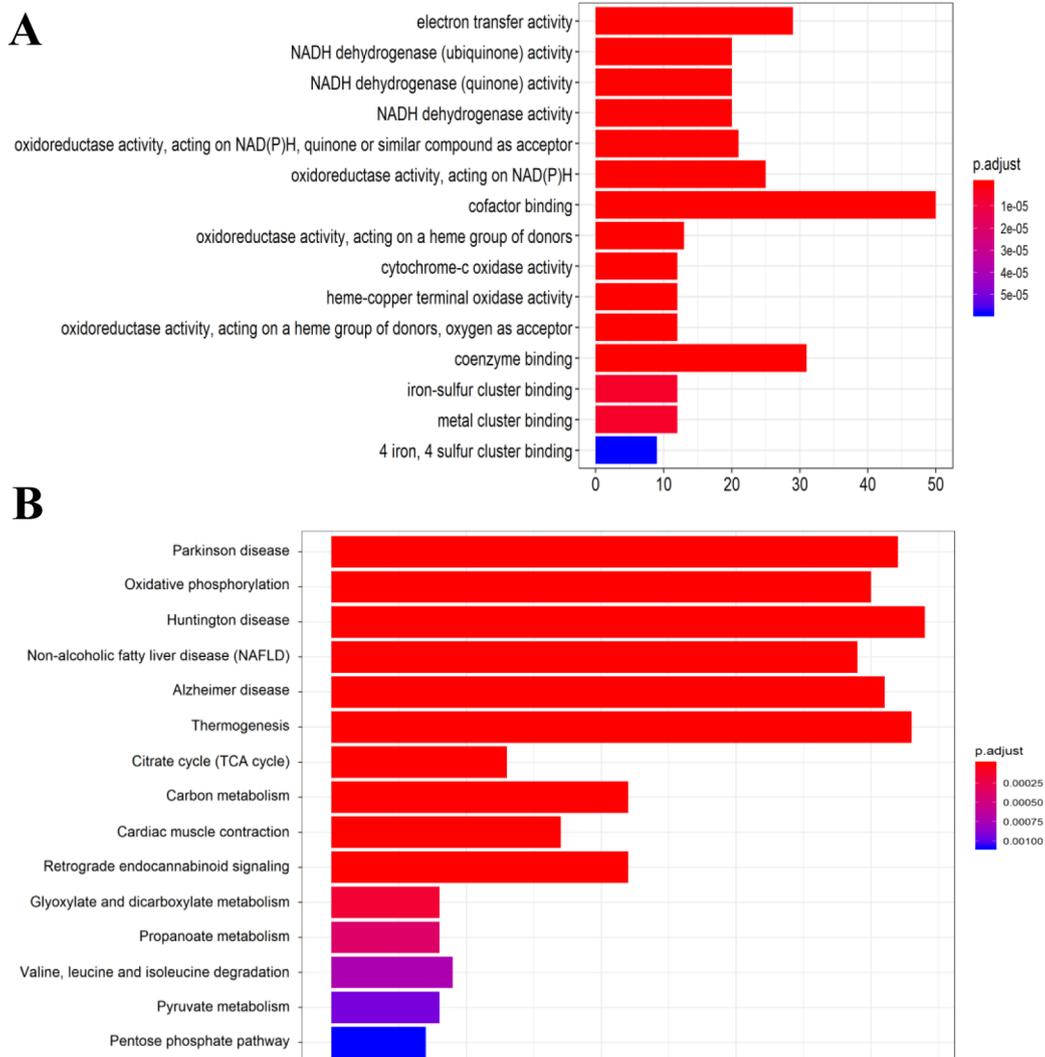


Figure 3. Enriched gene ontology (A) and Kyoto encyclopedia of genes and genomes (B) pathway analysis of co-expressed genes related to IDI2-AS1 and XIST.

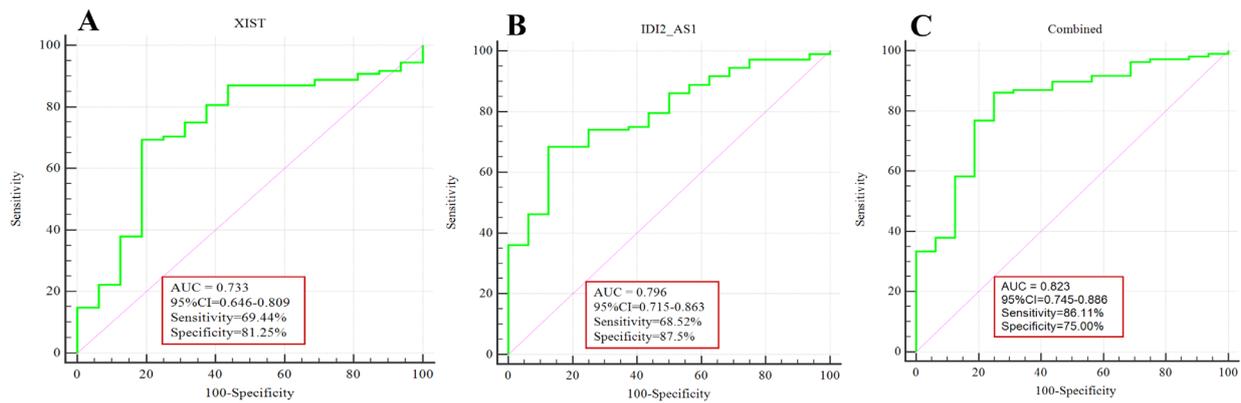


Figure 4. Diagnostic values of XIST (A), IDI2-AS1 (B), and IDI2-AS1 combined XIST (C) in discriminating dilated cardiomyopathy samples from non-failing control samples.

3.3. Diagnostic significance of differentially expressed lncRNAs

To explore the diagnostic significance of differentially expressed lncRNAs, ROC curves were generated. As shown in Figure 4, the diagnostic effectiveness of two lncRNAs in discriminating DCM from non-failing control ones revealed a favorable diagnostic value with an AUC of 0.733 (95% CI: 0.646-0.809) in XIST (Figure 4A) and AUC of 0.796 (95% CI: 0.715-0.863) in IDI2-AS1 (Figure 4B). Moreover, when XIST and IDI2-AS1 were combined, the ROC curve yielded an AUC of 0.823 (95% CI: 0.745-0.886, Figure 4C).

4. Discussion

DCM is the third most common cause of congestive heart failure and sudden cardiac arrest (17). Previous studies have revealed that lncRNAs play important roles in cardiac development and participate in cardiac homeostasis and regeneration as epigenetic regulators of heart gene expression (18,19). Recently, increasing attention has been paid to improving the treatment of DCM. However, there remains a long road to better understand of the underlying mechanism. In this study, dysregulated lncRNAs between patients with DCM and non-failing controls were analyzed by RNA sequencing. In total, a differentially expressed lncRNA steadily upregulated and another steadily downregulated. In particular, based on the lncRNA-related protein-coding genes, the following functional enrichment analysis indicated that lncRNA-related co-expressed mRNAs were mainly enriched in NADH activity, cytochrome c oxidase activity, and oxidoreductase activity. KEGG pathway enrichment showed that these genes were mainly involved in oxidative phosphorylation and cardiac muscle contraction.

In recent years, several studies have explored the functions and clinical significance of lncRNAs in HCM (20,21). A recent study investigated 14 lncRNAs deregulated in patients with non-end-stage

dilated hypokinetic ischemic cardiomyopathy, such as CDKN2B-AS1, EGOT, and H19 (22). A whole-transcriptome analysis of heart biopsy specimens from patients with DCM and healthy heart donors confirmed that myocardial infarction associated transcript was highly expressed in patients with DCM (23). To date, many lncRNAs have been revealed to be associated with myocardial pathophysiology, and research on the biological function and mechanism of lncRNAs has only begun and remained unclear (24,25). LncRNA X inactive-specific transcript (XIST) is indispensable for transcriptional silencing of the X chromosome in female mammals, which play vital role in the inactivation of the X chromosome (26). It has been widely used as an oncogene or a tumor suppressor in many malignancies (27). XIST has also been demonstrated to be remarkably upregulated in the hypertrophic hearts of mice induced by transaortic constriction and phenylephrine-treated rat cardiomyocytes (28). Furthermore, it was shown that XIST could inhibit myocardial cell proliferation and promote apoptosis. It was also reported that XIST contributed to tumor proliferation and suppressed tumor apoptosis in lung cancer and hepatocellular carcinoma (29,30). All these findings help to explain the potential contribution of XIST to DCM. As for isopentenyl-diphosphate delta isomerase 2 antisense RNA 1 (IDI2-AS1), its function in DCM has not been reported in the literature. Overexpression of IDI2-AS1 sensitized human cells to cell death in response to various stresses, such as ultraviolet irradiation, cycloheximide, hydrogen peroxide, and mercury II chloride (31). Therefore, its function should be explored in more studies.

Systolic dysfunction is the first hallmark pathophysiological sign of DCM (32). In the normal heart, contraction is synchronous within the myocardium, with a normal symmetric distribution of negative strain across the wall. Conversely, in the heart with DCM, the contraction pattern is obviously dyssynchronous (33). Therefore, heart contraction may be a vital indicator of DCM. In heart failure, deterioration of cytosolic Ca^{2+} and Na^+ handling hinders mitochondrial Ca^{2+} uptake

and the subsequent Krebs cycle-induced regeneration of the decreased forms of NADH and NADPH, leading to energetic deficit and oxidative stress (34). These results were consistent with the enrichment analysis that lncRNAs were mainly enriched in NADH and oxidoreductase activity and cardiac muscle contraction. This finding was confirmed by the fact that reduced sarcomere contractility can increase ventricular volumes to maintain cardiac output through the Frank-Starling mechanism, generating a thin-walled LV appearance that is discovered in DCM (35). DCM is usually diagnosed at ages between 30 and 40 years of age. However, owing to the vague symptoms and delay in diagnosis, the outcome for DCM is poor (36). DCM accounts for about one-third of heart failure patients. A recent study reported that lncRNAs ENST00000532365 and ENST00000507296 could serve as biomarkers for the diagnosis of DCM-related heart failure (36). In the present study, we found that XIST presented an AUC of 0.733, and IDI2-AS1 exhibited an AUC of 0.796. Moreover, when the two lncRNAs were combined, the ROC curve yielded an AUC of 0.823 with a sensitivity of 86.11% and specificity of 75.00%. Therefore, based on these results, XIST combined with IDI2-AS1 can be used as a novel diagnostic biomarker in DCM. However, the present study has some limitations. First, the sample size was relatively small. In addition, this study was performed based on bioinformatic methods, and the findings should be confirmed by experimental validation.

In summary, the present study identified IDI2-AS1 and XIST and related pathways that can help us understand the molecular mechanisms underlying the pathogenesis of DCM. XIST combined with IDI2-AS1 may be a novel promising novel biomarker for DCM diagnosis.

Data availability

The data that support the findings of this study are available from the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>), accession number GSE5406.

Acknowledgments

The authors thank the efforts of the National Cancer Institute in the creation of the Gene Expression Omnibus database

References

1. Elliott P. Cardiomyopathy. Diagnosis and management of dilated cardiomyopathy. *Heart*. 2000; 84:106-112.
2. Jefferies JL, Towbin JA. Dilated cardiomyopathy. *Lancet*. 2010; 375:752-762.
3. Hershberger RE, Hedges DJ, Morales A. Dilated cardiomyopathy: the complexity of a diverse genetic

- architecture. *Nat Rev Cardiol*. 2013; 10:531-547.
4. Felker GM, Thompson RE, Hare JM, Hruban RH, Clemetson DE, Howard DL, Baughman KL, Kasper EK. Underlying causes and long-term survival in patients with initially unexplained cardiomyopathy. *N Engl J Med*. 2000; 342:1077-1084.
5. Bozkurt B, Colvin M, Cook J, *et al*. Current Diagnostic and Treatment Strategies for Specific Dilated Cardiomyopathies: A Scientific Statement From the American Heart Association. *Circulation*. 2016; 134:e579-e646.
6. Thum T, Gross C, Fiedler J, *et al*. MicroRNA-21 contributes to myocardial disease by stimulating MAP kinase signalling in fibroblasts. *Nature*. 2008; 456:980-984.
7. Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics*. 2013; 193:651-669.
8. Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature*. 2011; 470:284-288.
9. Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet*. 2014; 15:7-21.
10. Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*. 2011; 147:358-369.
11. Ginger MR, Shore AN, Contreras A, Rijnkels M, Miller J, Gonzalez-Rimbau MF, Rosen JM. A noncoding RNA is a potential marker of cell fate during mammary gland development. *Proc Natl Acad Sci U S A*. 2006; 103:5781-5786.
12. Thum T, Condorelli G. Long noncoding RNAs and microRNAs in cardiovascular pathophysiology. *Circ Res*. 2015; 116:751-762.
13. Gomes CPC, Spencer H, Ford KL, Michel LYM, Baker AH, Emanuelli C, Balligand JL, Devaux Y. The function and therapeutic potential of long non-coding RNAs in cardiovascular development and disease. *Mol Ther Nucleic Acids*. 2017; 8:494-507.
14. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015; 43:e47.
15. Bhardwaj N, Lu H. Correlation between gene expression profiles and protein-protein interactions within and across genomes. *Bioinformatics*. 2005; 21:2730-2738.
16. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012; 16:284-287.
17. Halliday BP, Cleland JGF, Goldberger JJ, Prasad SK. Personalizing risk stratification for sudden death in dilated cardiomyopathy: The past, present, and future. *Circulation*. 2017; 136:215-231.
18. Devaux Y, Zangrando J, Schroen B, Creemers EE, Pedrazzini T, Chang CP, Dorn GW, 2nd, Thum T, Heymans S. Long noncoding RNAs in cardiac development and ageing. *Nat Rev Cardiol*. 2015; 12:415-425.
19. Calore M, De Windt LJ, Rampazzo A. Genetics meets epigenetics: Genetic variants that modulate noncoding RNA in cardiovascular diseases. *J Mol Cell Cardiol*. 2015; 89:27-34.
20. Sun J, Wang C. Long non-coding RNAs in cardiac

- hypertrophy. *Heart Fail Rev.* 2019.
21. Wang K, Long B, Zhou LY, Liu F, Zhou QY, Liu CY, Fan YY, Li PF. CARL lncRNA inhibits anoxia-induced mitochondrial fission and apoptosis in cardiomyocytes by impairing miR-539-dependent PHB2 downregulation. *Nat Commun.* 2014; 5:3596.
 22. Greco S, Zaccagnini G, Perfetti A, Fuschi P, Valaperta R, Voellenkle C, Castelvechio S, Gaetano C, Finato N, Beltrami AP, Menicanti L, Martelli F. Long noncoding RNA dysregulation in ischemic heart failure. *J Transl Med.* 2016; 14:183.
 23. Frade AF, Laugier L, Ferreira LR, *et al.* Myocardial infarction-associated transcript, a long noncoding RNA, is overexpressed during dilated cardiomyopathy due to chronic chagas disease. *J Infect Dis.* 2016; 214:161-165.
 24. Zhang J, Feng C, Song C, Ai B, Bai X, Liu Y, Li X, Zhao J, Shi S, Chen X, Su X, Li C. Identification and analysis of a key long non-coding RNAs (lncRNAs)-associated module reveal functional lncRNAs in cardiac hypertrophy. *J Cell Mol Med.* 2018; 22:892-903.
 25. Chen C, Tang Y, Sun H, Lin X, Jiang B. The roles of long noncoding RNAs in myocardial pathophysiology. *Biosci Rep.* 2019; 39:BSR20190966.
 26. Lo HC, Wu SC, Huang HC, Yeh CC, Huang JC, Hsieh CH. Laparoscopic simple closure alone is adequate for low risk patients with perforated peptic ulcer. *World J Surg.* 2011; 35:1873-1878.
 27. Yang Z, Jiang X, Jiang X, Zhao H. X-inactive-specific transcript: A long noncoding RNA with complex roles in human cancers. *Gene.* 2018; 679:28-35.
 28. Zhou T, Qin G, Yang L, Xiang D, Li S. LncRNA XIST regulates myocardial infarction by targeting miR-130a-3p. *J Cell Physiol.* 2019; 234:8659-8667.
 29. Mo Y, Lu Y, Wang P, Huang S, He L, Li D, Li F, Huang J, Lin X, Li X, Che S, Chen Q. Long non-coding RNA XIST promotes cell growth by regulating miR-139-5p/PDK1/AKT axis in hepatocellular carcinoma. *Tumour Biol.* 2017; 39:1010428317690999.
 30. Tang Y, He R, An J, Deng P, Huang L, Yang W. lncRNA XIST interacts with miR-140 to modulate lung cancer growth by targeting iASPP. *Oncol Rep.* 2017; 38:941-948.
 31. Tani H, Torimura M. Development of cytotoxicity-sensitive human cells using overexpression of long non-coding RNAs. *J Biosci Bioeng.* 2015; 119:604-608.
 32. Garfinkel AC, Seidman JG, Seidman CE. Genetic pathogenesis of hypertrophic and dilated cardiomyopathy. *Heart Fail Clin.* 2018; 14:139-146.
 33. Curry CW, Nelson GS, Wyman BT, Declerck J, Talbot M, Berger RD, McVeigh ER, Kass DA. Mechanical dyssynchrony in dilated cardiomyopathy with intraventricular conduction delay as depicted by 3D tagged magnetic resonance imaging. *Circulation.* 2000; 101:E2.
 34. Bertero E, Maaack C. Calcium signaling and reactive oxygen species in mitochondria. *Circ Res.* 2018; 122:1460-1478.
 35. Lakdawala NK, Thune JJ, Colan SD, Cirino AL, Farrohi F, Rivero J, McDonough B, Sparks E, Orav EJ, Seidman JG, Seidman CE, Ho CY. Subtle abnormalities in contractile function are an early manifestation of sarcomere mutations in dilated cardiomyopathy. *Circ Cardiovasc Genet.* 2012; 5:503-510.
 36. Zhang X, Nie X, Yuan S, Li H, Fan J, Li C, Sun Y, Zhao Y, Hou H, Wang DW, Chen C. Circulating long non-coding RNA ENST00000507296 is a prognostic indicator in patients with dilated cardiomyopathy. *Mol Ther Nucleic Acids.* 2019; 16:82-90.

Received February 14, 2020; Revised July 4, 2020; Accepted July 23, 2020.

[§]These authors contributed equally to this work.

*Address correspondence to:

Yushun Zhang, Department of Structural Heart Disease, The First Affiliated Hospital of Xi'an Jiaotong University, No.277, Yanta West Road, Xi'an 710061, China.

E-mail: zys2889@sina.com

Released online in J-STAGE as advance publication July 29, 2020.