ABSTRACT

Objective: Ischemic stroke and myocardial infarction are 2 of the leading causes of mortality. Both conditions are caused by arterial occlusion, resulting in ischemic necrosis of the cells in the cortex and heart. Long non-coding RNAs (lncRNAs) are a group of non-coding RNAs longer than 200 nucleotides without protein-coding potential. Thousands of lncRNAs have been identified but their involvement in ischemic stroke and myocardial infarction has not been studied extensively. Therefore, this study aimed to identify the role of lncRNAs, particularly those that are commonly altered in these two ischemic injuries.

Methods: We combined diverse RNA sequencing data obtained from public databases and performed extensive bioinformatics analyses to determine reliable lncRNAs commonly identified from these datasets. Using sequence analysis, we also detected the lncRNAs that may act as microRNA (miRNA) regulators.

Results: We found several altered lncRNAs that were common in ischemic stroke and myocardial infarction models. Some of these lncRNAs, including zinc finger NFX1-type containing 1 antisense RNA 1 and small nucleolar RNA host gene 1, were previously reported to be involved in the pathogenesis of each of these models. Interestingly, several lncRNAs had binding sites for miRNAs that were previously reported to be involved in the hypoxic response, suggesting the possible role of these lncRNAs as regulators in ischemic responses.

Conclusion: The lncRNAs identified in this study will be useful in determining the regulatory networks in ischemic stroke and myocardial infarction and in identifying potential specific markers for each of these ischemic diseases.

Keywords: Long non-coding RNA; Myocardial infarction; Stroke; MicroRNAs

INTRODUCTION

Ischemic stroke and myocardial infarction are 2 of the most common diseases that result in high mortality.1,2 Although the 2 conditions mainly differ in the speed of cell death and loss of function, their common pathophysiology lays in the acute occlusion of arteries, resulting in ischemic necrosis of the affected tissues. A frequently used molecular marker to diagnose myocardial infarction is troponin, released into circulation as a result of myocardial necrosis.3 However, the molecular marker for ischemic stroke has not yet been identified. Reperfusion
therapy is one of the most effective treatment methods for myocardial infarction, but the same method may cause a clinical problem during the treatment of ischemic stroke.\(^4\)

Long non-coding RNAs (lncRNAs) are a large group of non-coding RNAs with transcript length longer than 200 nucleotides,\(^5\) and diverse genomic locations. A proportion of lncRNAs is situated in the intergenic region between protein-coding genes—therefore called as long intergenic non-coding RNAs—whereas another proportion is localized near the protein-coding genes. The latter includes antisense lncRNAs that are positioned at the same genomic locus with protein-coding genes but with the opposite genomic direction. Moreover, another large population of lncRNAs is sited near the promoter region of protein-coding genes and produces transcripts in the direction opposite (divergent transcription) to that of protein-coding genes.\(^6\) The numbers of discovered lncRNAs in humans and mice are still growing and a recent GENCODE annotation reported the number of lncRNA genes (≤18,000) comparable to that of protein-coding genes (≤20,000) in humans.\(^7\) The primary working mechanism of lncRNAs includes associating with transcription factors to regulate the transcription of other genes and binding to microRNAs (miRNAs) to block their post-transcriptional regulation.\(^5\) Although the lncRNAs have been studied in diverse human diseases, their roles and working mechanism in myocardial infarction and ischemic stroke have not been investigated extensively. Compared to the protein-coding genes, for which many expression-profiling studies during the pathogenesis of these diseases have been conducted, few studies have been performed for the analysis of lncRNAs in the same models.

In this study, we ascertained the lncRNAs commonly altered in the 2 ischemic disease models: myocardial infarction and ischemic stroke. Based on publicly available data, we identified several lncRNAs that may function as important regulators in these diseases. Besides, we also showed that several lncRNAs may work as regulators of miRNAs. The information reported in this study will contribute to future research on the pathology of myocardial infarction and ischemic stroke.

**MATERIALS AND METHODS**

1. **Analysis of RNA sequencing data**

RNA sequencing data were obtained from the Gene Expression Omnibus (GEO) database.\(^8\) For myocardial infarction datasets, the raw FASTQ data of datasets with accession numbers GSE95759\(^9\) and GSE104187\(^10\) were downloaded. For the GSE95759 dataset, samples from adult mice were used (4 control and 4 infarcted samples, respectively), and for GSE104187, samples prepared 3 days after their surgery were used (2 control and 2 infarcted samples, respectively). For ischemic stroke datasets, FASTQ datasets with accession numbers GSE104882 and GSE112348\(^11\) were obtained. For the GSE104882 dataset, samples derived from the cortical tissues of mice 3 days after their surgery were used (3 control and 4 stroke samples, respectively). Only mice raised on a normal diet were selected for our analysis. In the case of the GSE112348 dataset, samples prepared 24 hours after their surgery were used for our analysis (3 control and 3 stroke samples, respectively). The procedure to analyze FASTQ data has been described in a previous study.\(^12\) Briefly, reads with low sequencing quality were removed using the Trimmomatic algorithm.\(^13\) The remaining sequences were aligned into the mouse genome (mm10) using STAR\(^14\) and the Fragments Per Kilobase of transcript per Million mapped reads (FPKM) value was calculated using Cuffnorm\(^15\) based
on the GENCODE annotation (version M17). Using the FPKM value for each gene, the fold change and \( p \)-value based on the \( t \)-test were calculated.

To increase the reliability of the results from the analysis algorithm, we also used another approach to calculate the fold ratio of gene expression between treatment and control groups in each dataset. Using the FASTQ reads after filtering out low-quality reads, the transcript level of each gene was analyzed using the Salmon algorithm\(^{16} \) based on the GENCODE annotation. Fold ratio and \( p \)-value of each gene were calculated using the edgeR software.\(^{17} \)

Among the analyzed genes, the significantly changed lncRNAs were selected, provided the lncRNAs were in the top 10% groups based on \( p \)-values in both analysis workflows, i.e. STAR-Cuffnorm and Salmon-edgeR. We combined the results from each of the myocardial infarction and ischemic stroke groups as described in Fig. 1B. To select the altered lncRNAs common between myocardial infarction and ischemic stroke models, we intersected the combined data between the 2 disease datasets. For miRNA target analysis, only the lncRNAs with FPKM values greater than 10 in both datasets were selected.

2. Prediction of miRNA target sites in lncRNA sequences

To identify a regulatory relationship between selected lncRNAs and miRNAs, we used the TargetScan algorithm to predict the possible binding sites of miRNAs in lncRNA sequences.\(^{18} \)

We only selected target sites containing a perfect 7-mer match of the seed sequence (thus, a perfect match of the 2nd to the 8th nucleotide from the 5′ end of the miRNA).

To increase the reliability of the prediction, public small RNA sequencing data for myocardial infarction (GSE79050) and ischemic stroke (GSE104037) were obtained from the GEO database.\(^{19,20} \) Among the miRNAs, only those with a significant difference between the control and treatment groups (\( p \)-value<0.05) were selected. By comparing the predicted miRNA targets from TargetScan analysis and the differentially expressed miRNAs identified

---

**Fig. 1.** Analysis of RNA sequencing data from MI and ischemic stroke (MCAO) models. (A) Changes in the expression of marker genes from each experimental model. The mRNA level change of Timp1 was used to determine the reliability of the MI model. For the MCAO model, the mRNA level of the Gfap gene, which encodes glial fibrillary acidic protein, was used as a marker. The \( p \)-value was calculated using Student’s \( t \)-test and described as asterisks. (B) The analysis scheme to identify altered lncRNAs common to MI and MCAO models. Details of the analysis procedure are described in the materials and methods section. Timp1, tissue inhibitor of metalloproteinase 1; Gfap, glial fibrillary acidic protein; MI, myocardial infarction; MCAO, middle cerebral artery occlusion; lncRNA, long non-coding RNA; mRNA, messenger RNA.

\(^{*}p<0.01, ^{†}p<0.001, ^{‡}p<0.0001.\)
from the analysis of public sequencing data, we shortlisted the miRNAs that had binding sites in lncRNAs and also showed anti-correlation in the direction of expression change to the same lncRNA.

RESULTS

1. Selection of altered lncRNAs common in myocardial infarction and ischemic stroke models

To identify the altered lncRNAs common in the injured heart from myocardial infarction and the injured cortex from ischemic stroke, we collected public RNA sequencing data available in the GEO database. For myocardial infarction data, we first selected GSE95755. In this dataset, myocardial infarction was induced by the permanent ligation of the left anterior descending coronary artery, and 3 days later, cardiomyocytes and fibroblasts were isolated for RNA sequencing. GSE104187 was the other dataset that we used for myocardial infarction. The samples used to make this dataset was also prepared 3 days after the ligation of the left anterior descending artery, but the RNA sequencing was performed for total heart tissue. In the case of the ischemic stroke model, GSE104882 and GSE112348 were used, wherein the RNA sequencing was performed using mouse cortex samples 3 days (GSE104882) and 1 day (GSE112348) after the middle cerebral artery occlusion (MCAO) operation, respectively. After the initial quality check of RNA sequencing data and quantitation of both protein-coding and non-coding genes, we checked whether the datasets that we chose were reliable, using marker gene levels. For the myocardial infarction model, tissue inhibitor of metalloproteinase 1 (Timp1) was selected as the marker gene. It was previously reported that Timp1 was significantly increased in myocardial infarction. In the case of the ischemic stroke model, we selected glial fibrillary acidic protein, that was reported as a reliable marker in stroke model. In the RNA sequencing data, these 2 marker genes were significantly increased, confirming the reliability of the datasets that we selected (Fig. 1A).

We selected differentially expressed lncRNAs in each dataset and combined the list of lncRNAs from the 2 myocardial infarction datasets and the 2 ischemic stroke datasets (Fig. 1B). The list of differentially expressed lncRNAs is included in Supplementary Table 1. By merging the combined myocardial infarction and ischemic stroke datasets, the altered lncRNAs common to both the experimental models were selected. These analyses resulted in the identification of one commonly decreased and 8 commonly increased lncRNAs in the 2 ischemic models (Fig. 2).

2. Genomic analysis of selected lncRNAs

The expression of the lncRNA RP23-445K23.4 was decreased significantly in both, myocardial infarction and ischemic stroke models (Fig. 2A). Interestingly, one of the isoforms of RP23-445K23.4 includes the sequence of miRNAs miR-29b-2 and miR-29c, suggesting this lncRNA works as the primary transcript to produce these miRNAs. A previous report showed that miR-29 family members were downregulated after myocardial infarction, and miR-29b levels were decreased following acute ischemic stroke. Thus, it is plausible that RP23-445K23.4 is involved in the progression of both these diseases by decreasing the production of miR-29 family miRNAs.

The lncRNA RP23-234K24.8 (also annotated as Lrrc75-as1)—significantly increased in both the ischemic models—is located downstream of the Lnc75a gene with the opposite direction

Fig. 2. Genomic information near the identified lncRNAs. For each lncRNA and its neighboring genes (A-I), the gene structures and genomic coordinates were obtained from the UCSC genome browser (http://genome.ucsc.edu/). For the datasets where significant changes in expression of lncRNA were observed, the level of expression change is depicted on the right. The p-value was calculated using Student’s t-test and described as asterisks. Each exon and intron is indicated with a filled box and solid line, respectively, and the gene orientation is indicated with an arrowhead. Genes written in blue indicate protein-coding genes, whereas genes written in green and red designate lncRNAs and small non-coding RNAs, respectively.

MI, myocardial infarction; MCAO, middle cerebral artery occlusion; lncRNA, long non-coding RNA; Zfas1, zinc finger NFX1-type containing 1 antisense RNA 1; Snhg, small nucleolar RNA host gene.

*p<0.05, †p<0.005, ‡p<0.0005.
of transcription (Fig. 2B). This lncRNA contains the sequences of several small nucleolar RNAs (snoRNAs) including snord49b, snord49a, and snord65. Our previous study showed that Lrrc75-as1 regulates vascular calcification in vascular smooth muscle cells. Another lncRNA Gas5 was also increased in both the models and contains several snoRNAs as its intronic sequences (Fig. 2C). It has been reported that Gas5 regulates myocardial infarction and ischemic stroke by targeting miR-525-5p and miR-437, respectively.27,28

The lncRNA zinc finger NFX1-type containing 1 (Znf1) antisense RNA 1 (Zfas1) was increased in both the models (Fig. 2D). The first exon of Zfas1 overlaps with that of protein-coding gene Znf1. Zfas1 was reported to be increased in a mouse myocardial infarction model and operates as a Serca2a inhibitor by binding to this protein.29 In contrast, this lncRNA was downregulated in the blood leukocytes of patients with ischemic stroke,30 requiring the measurement of its levels in the cortex from the ischemic stroke model. Small nucleolar RNA host gene (Snhg) 1 is the host gene of many snoRNAs and was increased in both the models (Fig. 2E). This lncRNA has been reported to regulate cerebrovascular pathology in the ischemic stroke model by regulating HIF-1α/VEGF signaling,31 although there is no report indicating the role of Snhg1 in myocardial infarction until now.

The lncRNAs RP23-349B4.7, RP23-110C17.2, and Snhg5 were increased in myocardial infarction and ischemic stroke models (Fig. 2F-H). Among them, RP23-349B4.7 and Snhg5 contain snoRNA in their intronic sequences (Fig. 2F and H). H19 is a well-known lncRNA that contains miR-675 in its sequence (Fig. 2I). The expression of this lncRNA is dysregulated in diverse cancers and its working mechanism had been extensively studied in cancer models.32 Compared to the many studies in cancer, few studies have been reported in myocardial infarction and ischemic stroke for H19 lncRNA. There is a report that H19 was increased in the rat model of cerebral ischemia, and it promoted neuroinflammation by driving histone deacetylase 1-dependent M1 microglial polarization.33 However, another study reported that H19 was decreased in infarcted myocardium.34 In our analysis, the expression of H19 was increased in both the models (Fig. 2I). Because myocardial infarction and ischemic stroke have many common phenotypes including hypoxic responses, the expression of H19 is expected to show the same trend between 2 models. Thus, we support the data that H19 is increased in both ischemic models.

3. Identification of miRNA targets of lncRNAs

To predict the functions of the common lncRNAs, we selected the lncRNAs with high expression in samples from both the models. One of the main working mechanisms of lncRNAs is the inhibition of miRNA function through sequence-specific binding.3 To operate as an efficient miRNA suppressor, the lncRNAs need to be expressed at a high level as described previously.35 We arbitrarily selected the expression cutoff value of FPKM as 10, and the selected lncRNAs based on this cutoff included RP23-234K24.8, Gas5, Zfas1, Snhg1, and RP23-349B4.7, all of which were upregulated in both the models. To select the miRNAs that can base-pair with the lncRNAs, we used the TargetScan algorithm to predict the binding between miRNAs and lncRNAs and organized miRNA-lncRNA pairs (Fig. 3A).

LncRNAs and miRNAs suppress each other, thereby showing anti-correlation in their expression. To select reliable lncRNA-miRNA pairs, we obtained the small RNA sequencing data from public databases. The dataset with the accession number GSE79050 represents data from a myocardial infarction model, while GSE104037 denotes data from an MCAO model.19,20 After choosing the miRNAs that were differentially expressed between untreated
and treated samples, the miRNAs with expression change in the direction opposite to that of each lncRNA were selected. Among the selected miRNA-lncRNA pairs, those that contained base-pairing sequences predicted by TargetScan were chosen (Fig. 3A). The resultant lncRNA-miRNA pairs included Zfas1-miR-136-5p, Snhg1-let-7i-5p, Snhg1-miR-141-3p, Snhg5-miR-448-3p, and Snhg5-miR-141-3p (Fig. 3B-D). The expression of these miRNAs was decreased in the treatment samples in each dataset, as shown in Fig. 3E.

**Fig. 3.** Analysis of lncRNA-targeting miRNA. (A) Analysis scheme to identify miRNA targets of lncRNAs. The miRNA targets were identified by combining the results from in silico predictions of binding sequences in lncRNAs and those from analyzing the expression changes of miRNAs in small RNA sequencing data of MI and ischemic stroke (MCAO) models. The detailed procedure is described in the Materials and Methods section. (B-D) The position and sequence of the miRNA-binding region in the lncRNA (B) Zfas1, (C) Snhg1, and (D) Snhg5. The sequence matches between miRNA and lncRNA were predicted using the mfold web server (http://unafold.rna.albany.edu/?q=mfold). (E) The expression levels of selected miRNAs in (B-D). In each dataset of MI and MCAO, the normalized counts of each miRNA between sham and treated groups are shown. *p*-value calculated using Student’s *t*-test is shown.

miRNA, microRNA; lncRNA, long non-coding RNA; FPKM, Fragments Per Kilobase of transcript per Million mapped reads; Zfas1, zinc finger NFX1-type containing 1 antisense RNA 1; Snhg, small nucleolar RNA host gene; MI, myocardial infarction; MCAO, middle cerebral artery occlusion.
DISCUSSION

In this study, we identified altered lncRNAs common between 2 ischemic disease models: myocardial infarction and ischemic stroke. Among the selected lncRNAs in this study, several lncRNAs including Gas5 were reported to be involved in myocardial infarction or ischemic stroke as described above. Because the expression of these lncRNAs showed the same pattern of expression change, it can be expected that they may be involved in physiological processes such as hypoxic response or cell death, common to these diseases. Moreover, 8 out of 9 selected lncRNAs are conserved in the human genome, suggesting they could have regulatory roles during the pathogenesis of human diseases (Fig. 2 and Supplementary Fig. 1). It will be interesting to uncover the detailed working mechanisms of these lncRNAs.

Although the lncRNAs common to both, myocardial infarction and ischemic stroke models were the primary targets of our analysis, many lncRNAs were altered in just one disease model exclusively (Supplementary Table 1). Thus, it is possible that these lncRNAs are involved in the processes occurring in only one disease model. The list of these lncRNAs will be a valuable resource for researchers aiming to determine the working mechanism of each of these ischemic diseases. However, we also note that the expression changes of lncRNAs after cerebral or myocardial ischemia could be affected by the selection of time points after the surgery. Thus, the measurement of selected lncRNAs at diverse time points after the surgery will be an appropriate starting point before analyzing the lncRNAs function.

For the lncRNAs, including Zfas1, Snhg1, and Snhg5, we identified several miRNAs that may be in a regulatory relationship with lncRNAs. Among those miRNAs, miR-141-3p has been shown to regulate hypoxia-induced apoptosis in cardiomyocytes.37 Because miR-141-3p was decreased in myocardial infarction-induced samples, and miR-141-3p-targeting lncRNAs—Snhg1 and Snhg5—were increased in myocardial infarction cases from our analyses, further studies are required to verify whether these lncRNAs are also involved in hypoxia-related processes via regulation of miR-141-3p.

Along with the research to elucidate the working mechanisms of lncRNAs or their-related miRNAs selected above, studies to discover molecular markers of these diseases are also essential. The list of overlapping or specifically changed lncRNAs in our study (Supplementary Table 1) is expected to be beneficial for studies aiming to discover the markers of myocardial infarction and ischemic stroke.

SUPPLEMENTARY MATERIALS

Supplementary Table 1
List of differentially expressed long non-coding RNAs in each dataset. The average signal means the average value of Fragments Per Kilobase of transcript per Million mapped reads of analyzed samples in each dataset. The fold change and p-value based on the t-test are included.
Supplementary Fig. 1
Genomic information near the human lncRNAs homologous to the selected mouse lncRNAs. For the lncRNAs in Fig. 2 except for RP23-110C17.2, which is not conserved in humans, their human homologs (hg19) were identified and depicted.

Click here to view

REFERENCES


