BBA - Molecular Basis of Disease 1864 (2018) 2266–2273

Contents lists available at ScienceDirect

BBA - Molecular Basis of Disease

journal homepage: www.elsevier.com/locate/bbadis

Systematic identification of rabbit LncRNAs reveals functional roles in atherosclerosis☆

Jia Li[a,b,c], Qianlan Yao[d], Fangyoumin Feng[a,b,c], Sheng He[a,b,c], Ping Lin[b,c], Liguang Yang[b,c], Chuhua Yang[b,c], Hong Li[b,*], Yixue Li[a,b,d,**]

[a] School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China
[b] CAS Key Laboratory for Computational Biology, CAS-MPG Partner Institute for Computing Biology, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[c] University of the Chinese Academy of Sciences, Beijing 100049, China
[d] School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200031, China

ARTICLE INFO

Keywords:
lncRNA
Rabbit
Atherosclerosis

ABSTRACT

Long noncoding RNAs (lncRNAs) have been gradually emerging as important regulators in various biological processes and diseases, while the contributions of lncRNAs to atherosclerosis remain largely unknown. Our previous work has discovered atherosclerosis associated protein-coding genes by transcriptome sequencing of rabbit models. Here we investigated the roles of lncRNAs in atherosclerosis. We defined a stringent set of 3736 multi-exonic lncRNA transcripts in rabbits. All lncRNAs are firstly reported and 609 (16.3%) of them are conserved in 13 species. Rabbit lncRNAs have similar characteristics to lncRNAs in other mammals, such as relatively short length, low expression, and highly tissue-specificity. The integrative analysis of lncRNAs and co-expressed genes characterize diverse functions of lncRNAs. Comparing two kinds of atherosclerosis models (LDLR-deficient WHHL rabbits and cholesterol-fed NZW rabbits) with their corresponding controls, we found the expression changes of two rabbit models were similar in aorta in but different in liver. The shared change in aorta revealed a subset of lncRNAs involved in immune response, while the cholesterol-fed NZW rabbits showed broader lncRNA expression changes in skeletal muscle system compared to WHHL rabbits. These atherosclerosis-associated lncRNAs and genes provide hits for the experimental validation of lncRNA functions. In summary, our study systematically identified rabbit lncRNAs for the first time and provides new insights for understanding the functions of lncRNAs in atherosclerosis. This article is part of a Special Issue entitled: Accelerating Precision Medicine through Genetic and Genomic Big Data Analysis edited by Yudong Cai & Tao Huang.

1. Introduction

Long noncoding RNAs (LncRNAs) are non-protein coding transcripts longer than 200 nucleotides. In recent years, more and more lncRNAs of model organisms have been identified by high-throughput sequencing. The structural, expression and evolutional characteristics of lncRNA have been widely studied in model organisms. A human lncRNA study showed that the overall transcript length of lncRNAs are smaller than coding genes while the exon and intron length of lncRNAs are longer [1]. LncRNAs express in a more tissue-specific manner than coding genes [2]. The evolutionary study of 17 species revealed that homologous lncRNAs share short, 5′-biased conserved patches while corresponding gene structures evolve rapidly [3]. However, lncRNAs in many species are not well defined, and the annotation of lncRNA functions still face big challenges.

Atherosclerosis is one of the most common vascular disorders in human. It is a systematic disease in which fatty deposits, immune responses, proliferation and apoptosis happened gradually inside the walls of arteries [4]. Although the exact cause of atherosclerosis is still a mystery, the cholesterol origin hypothesis had been proposed > 100 years ago [5], and inflammatory response is regarded as a driver in recent researches [6]. Rabbit is an important experimental animal for studying human hypercholesterolemia and atherosclerosis, because its lipid metabolism exhibits similar features with human [7]. Many important discoveries related to atherosclerosis were originated from rabbit models. New Zealand White (NZW) rabbit is the most widely

☆ This article is part of a Special Issue entitled: Accelerating Precision Medicine through Genetic and Genomic Big Data Analysis edited by Yudong Cai & Tao Huang.

* Corresponding author.
** Correspondence to: Y. Li, School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China.
E-mail addresses: lihong01@sibs.ac.cn (H. Li), yxli@sibs.ac.cn (Y. Li).

https://doi.org/10.1016/j.bbadis.2017.12.040
Received 16 October 2017; Received in revised form 4 December 2017; Accepted 28 December 2017
Available online 06 January 2018

used rabbit model. Cholesterol-fed NZW rabbits develop hypercholesterolemia and atherosclerosis, while cholesterol-fed mice did not. Watanabe heritable hyperlipidemic (WHHL) rabbit is an animal model of familial hypercholesterolemia, produced by selectively breeding of Japanese White (JW) rabbits [8]. Due to LDLR deletion, WHHL rabbits with normal diet have hypercholesterolemia and atherosclerosis.

Our previous work found some genetic modifiers in the pathophysiology of WHHL rabbits by whole-genome sequencing of two kinds of atherosclerosis rabbit models (cholesterol-fed NZW, WHHL) and the corresponding wild-type control (NZW, JW). We also found critical coding genes in atherosclerosis by transcriptome sequencing of multiple tissues of rabbit models, but the roles of noncoding transcripts are ignored [9]. A few of lncRNAs have been proved to play important roles in atherosclerosis. For example, lincRNA-p21 is a key regulator in atherosclerosis which can repress cell proliferation and induce programmed cell death in vascular smooth muscle cells and macrophage cells by enhancing p53 activity [10]. Lnc-Spry1 is identified as an immediate-early regulator of TGF-β-induced epithelial-mesenchymal transition which is one of the most curial transformations during the early stage of atherosclerosis [11]. Besides, previous studies linked lncRNA ANRIL located at human chromosome 9p21.3 with increased coronary artery disease risk [12]. Compared to these scattered studies, RNA-seq technology provides an easier way to identify new lncRNAs and explore their potential functions.

Here we systematically investigated 3736 rabbit lncRNAs by using the previously sequenced 69 samples. Next, we characterized each lncRNA through diverse features including transcript structure, expression pattern, evolutionary conservation, and putative functions from lncRNA-gene expression correlations. Furthermore, we investigated the expression changes of lncRNAs in two kinds of atherosclerosis rabbit models (cholesterol-fed NZW, WHHL), and found the common and specific roles of lncRNAs in two models.

2. Results

2.1. Identification of rabbit lncRNAs

We used our previous sequenced 69 rabbit samples to identify rabbit lncRNAs. These 69 samples represented six different tissue types, two kinds of atherosclerosis rabbit models and the corresponding wild-type controls. RNA sequencing data of each sample was mapped to rabbit genome and assembled to transcripts. Detailed information about the sample origin, mapping results and assembled transcripts are listed in Supplementary Table 1. Totally, 130,233 transcripts were produced from 69 samples. Protein-coding transcripts were selected by Cufflinks class code '=' or 'j'. This resulted in 97,280 protein-coding transcripts from 12,845 genes.

We developed a computational pipeline to infer multi-exonic lncRNAs from RNA sequencing data (Fig. 1A). Firstly, transcripts with Cufflinks class code 'i' or 'u' were chosen as candidate lncRNAs. Secondly, transcripts < 200 nt or having a single exon were filtered. Thirdly, coding potential of the remaining transcripts was scored by two tools: CPC [13] and PhyloCSF [14] (Method). Finally, the putative noncoding transcripts were translated in six reading frames, and scanned by HMMER to exclude transcripts which may encode any protein domain in Pfam [15] (E-value < 10^{-4}). In total, we identified 3739 lncRNA transcripts, corresponding to 1737 lncRNA genes with our pipeline. This firstly provided a comprehensive lncRNA identification from large-scale experimental data.

To remove accidently occurred transcripts, we only kept transcripts that were expressed in at least two samples in each tissue. This resulted in a final set of 3736 lncRNA transcripts from 1734 genes listed in Supplementary Table 2 and 89,820 coding transcripts from 12,480 genes. The 1734 lncRNA genes were classified into 343 genic (overlapped) and 1391 intergenic (455 divergent to coding genes) lncRNAs (Method). There were 1435 (11,430), 1318 (10,844), 1297 (11,140), 1277 (11,507), 1253 (10,815), 1201 (10,665) lncRNAs (genes) expressed in kidney, liver, aorta, embryo, heart and coronary. Principal component analysis based on the expression profiles of lncRNAs and protein-coding genes was conducted to reflect sample relationships and transcriptional level similarities (Fig. 1B). Samples from the same tissue were clustered together, except heart and coronary whose histological positions were close. Therefore, tissue origin had bigger effects on transcriptome than rabbit breed or disease state. Furthermore, different tissues were compared to show overlapped and specific LncRNA genes (Fig. 1C). Nearly half (891 out of 1734) lncRNA genes expressed in all tissues. 89.4% (1550) lncRNAs expressed in more than one tissues. 10.6% (184) lncRNA particularly expressed in one tissue. The same analysis was done for protein-coding genes (Supplemental Fig. 1). The percentages of overlapped genes in multiple-tissues are similar between lncRNAs and coding genes.

2.2. Structure characteristics of rabbit LncRNAs

To illustrate the structure characteristics of rabbit LncRNAs, transcript length, exon length, intron length and the number of isoforms were compared between LncRNAs and protein-coding genes (Fig. 2A, B, C). The length of LncRNA transcripts are one-third of the length of coding gene transcripts ($P < 2.2e\text{-}16$). On the contrary, lncRNA exons are extensively longer than the exons of coding genes ($P < 2.2e\text{-}16$). Moreover, lncRNA introns are also extensively longer than those from coding genes ($P < 2.2e\text{-}16$). These opposite conclusions mainly affected by the number of exons, because most lncRNA transcripts occupy less number of exons compared to coding gene transcripts (Supplemental Fig. 2). In addition, > 67% coding genes have at least 2 isoforms while the corresponding percentage of lncRNA genes is 38% (Fig. 2C). Collectively, the basic structure features of rabbit lncRNAs are consistent with other species [1, 16].

2.3. Tissue specificity of rabbit lncRNAs

To assess the tissue-specificity of lncRNAs, tissue-specificity index was calculated for each gene locus. Larger value represents higher tissue specificity. LncRNAs exhibit higher tissue-specific specificity than coding genes in all 6 tissues (Fig. 2D), consistent with previous observation [17]. Moreover, expression values in each sample were converted to the percentage across 69 samples. The clustering of lncRNAs expression got several obvious modules, presenting specific high-expressed lncRNAs in different tissues (Fig. 2E). In contrast, modules in the coding gene clustering were not as obvious as lncRNAs (Fig. 2E).

2.4. Evolutionary conservation of rabbit lncRNAs

Highly conservation could represent strong evidences for potential functions and provides a new way for further validation of function through homologous sequences from other species. To assess the evolutionary conservation of rabbit lncRNAs, lncRNAs of another 16 species were collected from NONCODE database [18] and were made as a reference library. By applying Blastn to rabbit lncRNA transcripts, 16.3% (n = 609) transcripts have homologous sequences in 13 species with a *P*-value cutoff 1e-4. In particular, 72.1% (n = 440) rabbit lncRNAs had homologous sequences in human (Fig. 3A). The conserved regions are usually short (median 230 bp), occupying ~18.4% of the whole transcripts. The majority (69%) of them have shown obvious 5′ bias, with the distance from the middle of the conserved part to 5′ end being shorter than to 3′ end. These characteristics are consistent with the previous observation in 17 species [3]. The majority of transcripts (50.1%, n = 306) only had homology in one species (Fig. 3B). Since the known lncRNAs are very incomplete, these observations need to be verified in future studies. We noted that one lncRNA (XLOC_003375) had two extremely conserved transcripts (TCONS_00013836 and TCONS_00013837). Both isoforms not only showed sequence homology

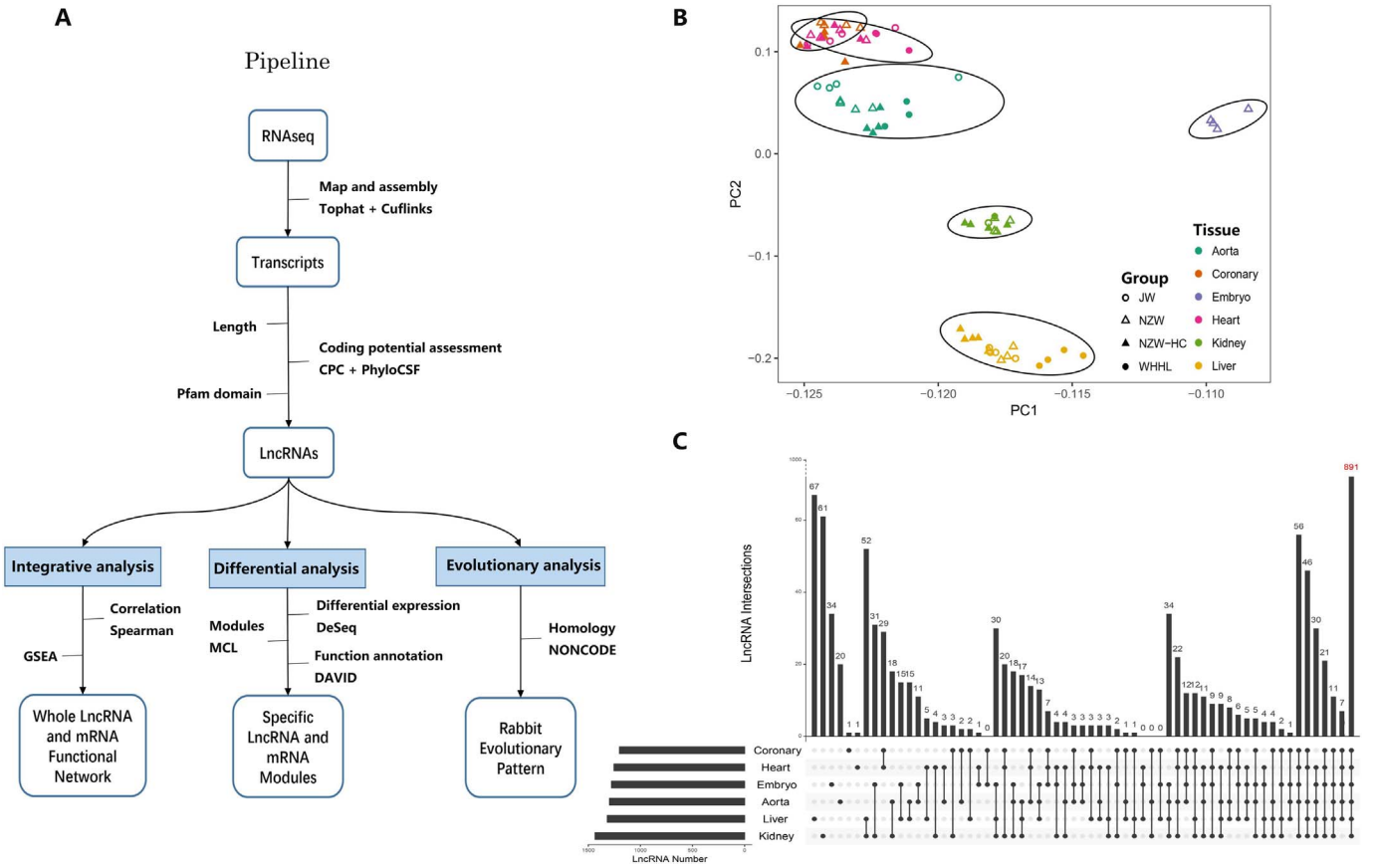

Fig. 1. Overview of RNA-seq based rabbit lncRNAs. (A) An integrative computational pipeline to analyze rabbit lncRNAs including identification, functional annotation and evolutionary analysis. (B) Principal component analysis based on all lncRNAs and coding genes. The first (PC1) and second component (PC2) are shown in the plot. The ellipses indicate samples with similar transcriptional level. Shapes and colors represent different experimental groups and tissue types, respectively. (C) An overview of overlapped and specific lncRNA genes in different tissues. Each vertical line represents a combination of multiple tissues. Each black point represents a tissue.

with 8 species (human, gorilla, orangutan, rhesus, mouse, cow, opossum and platypus), but also share the approximately 110 bp conservation patch within the all targeted homologous sequences (Fig. 3C). It locates in chromosome 11. The upstream gene RASA1 is a RAS activator influencing cell proliferation; downstream gene PLOR3G is a subunit of RNA polymerase III controlling DNA transcription. Combing the positon information with its functional annotation by GSEA (Method), this special lncRNA may be involved in the process of mitochondrial respiratory chain complex I assembly.

As there are little annotations for rabbit lncRNAs, we used lncRNAdb [19] database to investigate the functions of their homologous lncRNAs. Sixteen rabbit lncRNAs had homologous sequences in lncRNAdb. For example, a rabbit lncRNA was mapped to human ANRIL, which was annotated to be coronary heart disease related. However, the current available mouse lncRNAs did not have homologs of human ANRIL [12]. This result indicated that rabbit models are very valuable to discover atherosclerosis related lncRNAs.

2.5. Allocating function by expression correlation for all lncRNAs

Functional analysis for lncRNAs is always a hard challenge due to the lack of annotated information. At present, the universally used methods for lncRNA functional prediction are based on the expression correlation between lncRNAs and coding genes [2, 20]. Spearman correlation coefficients (SCC) were calculated for all lncRNA and coding genes [21]. For each lncRNA, GSEA enriched functions of its correlated genes were regarded as the potential functions of lncRNA. The majority of rabbit lncRNAs widely participate in diverse biological functions (Fig. 4A). Clustering analysis revealed several groups of biological processes that were involved in different lncRNAs [22]. Three dominant clusters were respectively visualized by a network of biological processes. Cluster A contained lipid metabolism and transport (Fig. 4B). Cluster E contained immune response, apoptosis, calcification, respiration and ion transport (Fig. 4C). Cluster G involved in heart muscle contraction (Fig. 4D).

2.6. Differential expression analysis in atherosclerosis

To investigate the functions of lncRNAs in atherosclerosis, we compared the difference between the atherosclerosis rabbit models and their corresponding wild-type controls. Specifically, cholesterol-fed NZW rabbits were compared with normal-fed NZW rabbits, and LDLR-deficient WHHL rabbits were compared with JW rabbits, respectively. The number of differentially expressed lncRNAs and genes in four tissues (aorta, heart, liver and kidney) were shown in Supplemental Fig. 3. A heatmap of differentially expressed lncRNAs illustrated the expression pattern in two atherosclerosis models and different tissues (Fig. 5A).

Heart and kidney did not have obvious expression change therefore they were removed in the further analysis. Aorta and liver were the major tissues affected by atherosclerosis. Compared to the LDLR-deficient WHHL models, cholesterol-fed NZW atherosclerosis models had more differential lncRNAs and genes in aorta and liver. This indicated that diet had more widely effects than monogenetic deficiency although both factors can cause atherosclerosis. Moreover, considering that aorta differential expressed lncRNAs had larger number (66 and 109 for WHHL and cholesterol-fed NZW) and higher overlapping rate (22.3%, n = 39) than those of liver (11 and 42 for WHHL and cholesterol-fed

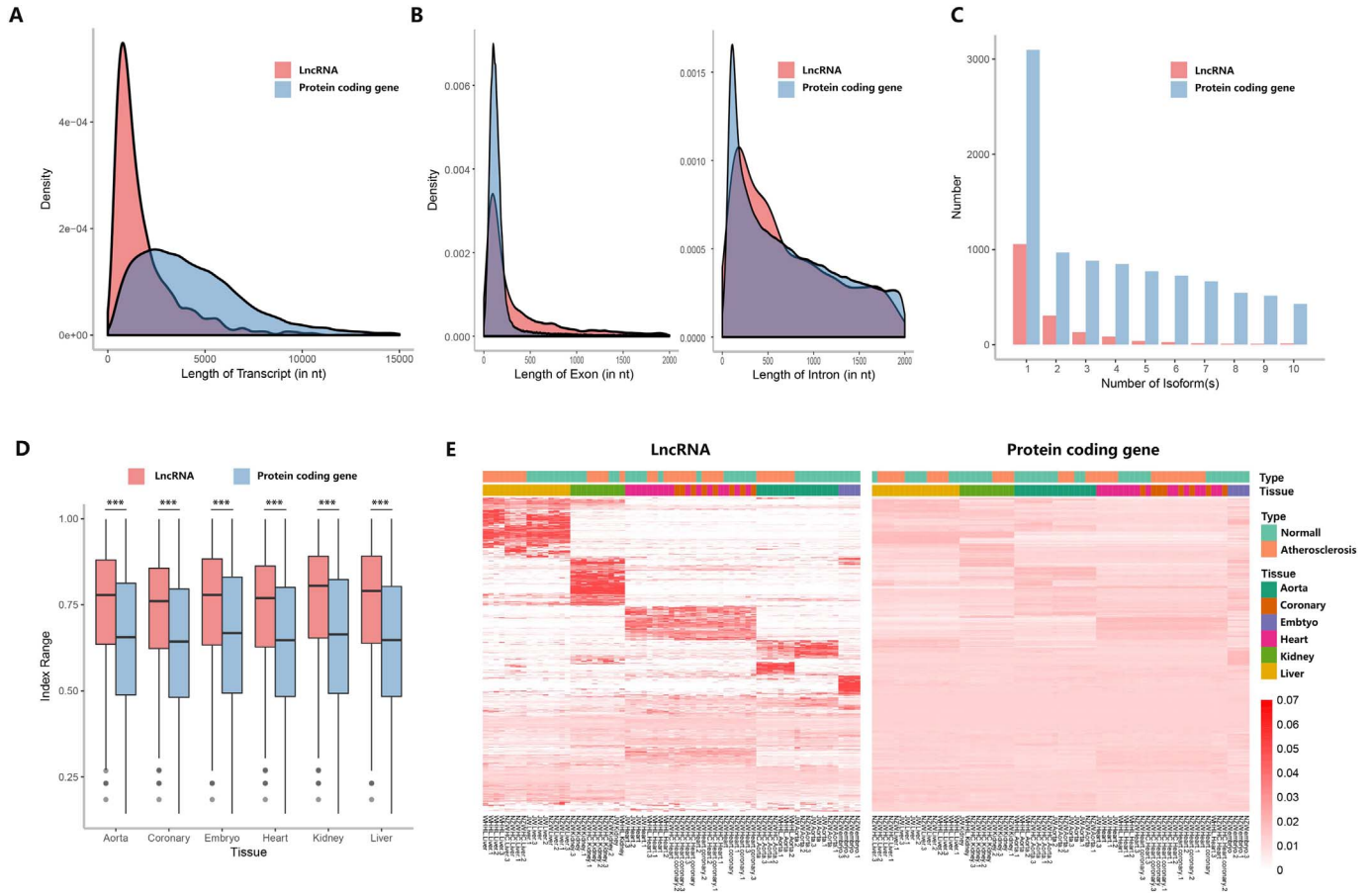

Fig. 2. Structure and expression characteristics of rabbit lncRNAs. (A) Size distribution of the processed transcripts. LncRNAs are generally shorter than protein coding transcripts. (B) Size distribution of exon (left) and intron (right). Both lncRNA exons and introns are longer than those of protein coding transcripts. (C) Comparison of the number of isoforms per lncRNA and protein-coding gene locus. (D) Distribution of tissue-specific index scores for lncRNAs and protein coding loci across different tissues. LncRNA loci display higher tissue specificity than protein coding loci in all 6 tissue types. Three asterisks (***) represent *P* value < .001. (E) Expression profiles of lncRNAs (left) and protein coding genes (right) across 6 tissue types and 2 disease situations. Color intensity represents the fraction of log-normalized FPKM expression values (see Methods section). Column name represents each sample identifier.

NZW; 0.019%, n = 1), we concentrated more on the aorta differential expressed lncRNAs during the further analysis. As the main disease occurrence place, aorta experienced a series of lesions starting from the accumulation of lipoprotein, immune response, smooth muscle cell migration, macrophage apoptosis to the finally thrombus formation [23]. We studied differentially expressed lncRNAs and coding genes in order to explain those unnoticed regulations happened together or alternatively between LDLR-deficient WHHL and diet influenced NZW rabbits as a reference for human researches with the same causes. Putative functions of those lncRNAs were assigned by the enrichment of correlated differential expressed coding genes. Cardiac muscle contraction and muscle filament sliding were significantly enriched in both models. Besides, phagosome acidification and phosphatidylinositol-mediated signaling were specific for cholesterol-fed NZW model while CD4-positive T cell differentiation was significantly alone for WHHL model (Fig. 5B).

2.7. Functional modules in aorta of atherosclerosis

Markov clustering algorithm (MCL) was used to recognize highly inter-connected functional modules including differentially regulated lncRNAs and coding genes in aorta. Seven and eight functional modules were identified for WHHL vs JW pair and cholesterol-fed NZW vs normal-fed NZW pair, respectively. The biggest module in cholesterol-fed NZW models contained 23,230 co-expressed pairs between 57 lncRNAs and 1470 genes, and the biggest module in WHHL models contained 3746 co-expressed pairs between 27 lncRNAs and 477 genes. The putative functions of each module were assigned based on the enriched pathways of coding genes. Then we investigated the shared and specific functional changes between two kinds of atherosclerosis models.

We compared the number of shared lncRNAs and genes between the top 3 dominant modules (Fig. 5C). The biggest modules of two atherosclerosis models were the most similar, which shared 8 lncRNAs and 424 genes. Moreover, they were also similar to each other in the pathway level, which were enriched for immune responses, such as regulation of leukocyte differentiation and activation especially lymphocyte (Supplemental Fig. 4A). In order to further determine the roles of 8 shared lncRNAs in atherosclerosis, their correlated coding genes were selected to do the Ingenuity Pathway Analysis (IPA). The 3 most significant pathways were shown in a lncRNA-gene correlation network (Fig. 5D, Supplemental Table 3), involved in Fcγ receptor-mediated phagocytosis in macrophages, NFAT regulation in immune response, and PI3k signal in B cells. Interestingly, the vast majority lncRNAs and coding genes were upregulated in two kinds of atherosclerosis models, in agreement with the previous assumption that immune responses were highly activated in atherosclerosis. Collectively, enrichment analysis for aorta lncRNAs regardless of disease causes suggested their major regulation functions on immune responses.

Next, we investigated the specific functional changes in two kinds of atherosclerosis models. We noticed that module 2 or 3 of both models enriched in skeletal muscle system (Supplemental Fig. 4B), but the detailed roles were different in cholesterol-fed NZW and WHHL. Muscles are composed of two major protein filaments: myosin and

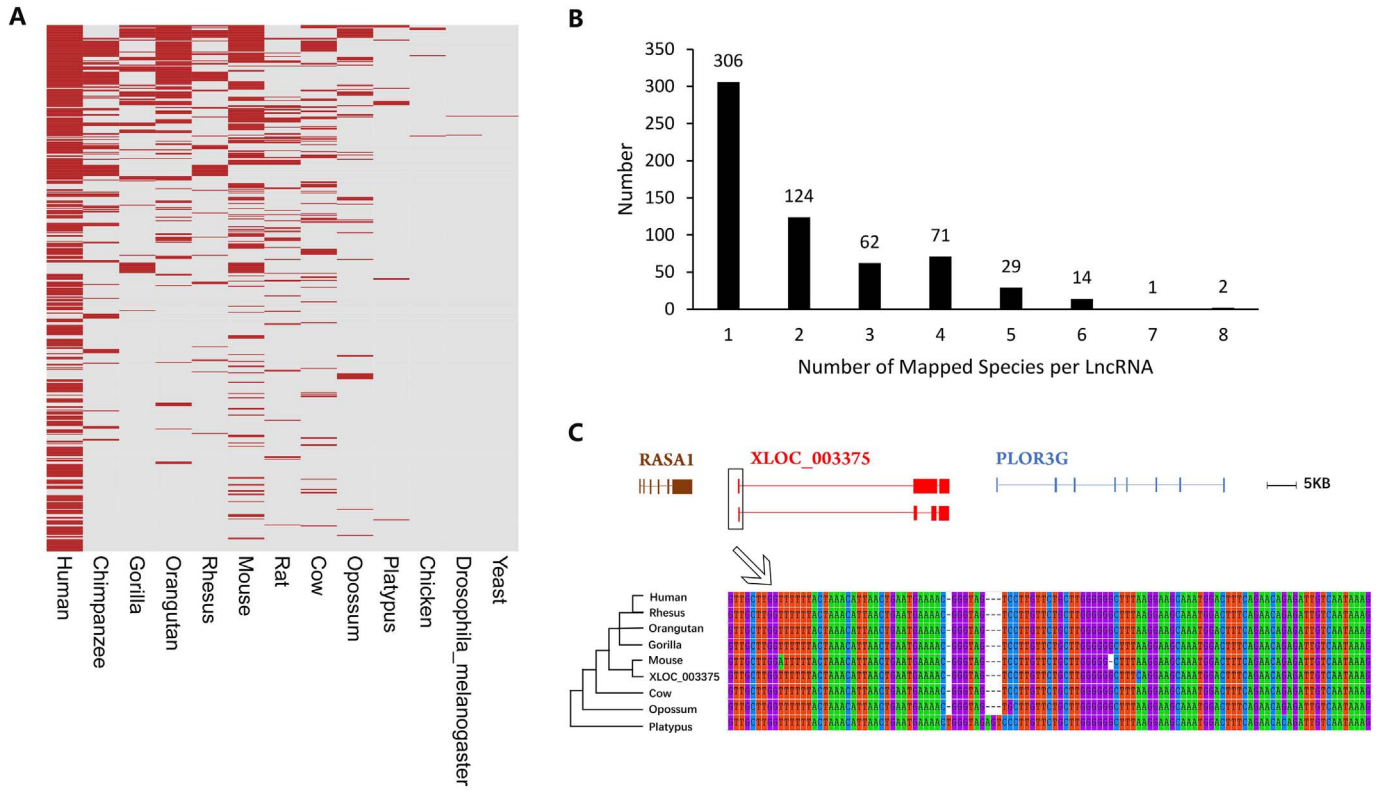

Fig. 3. Evolutionary conservation of rabbit lncRNAs. (A) Rabbit lncRNA conservation in 13 mammalian genomes. Red indicates the appearance of reliable homolog. (B) The number of homologs detected for each lncRNA. (C) A multiple sequence alignment of the lncRNA XLOC_003375 in mammal. Representative isoforms of lncRNA XLOC_003375 and corresponding upstream and downstream coding genes are shown.

actin. For the cholesterol-fed NZW model, 22 coding genes in module 2 participated in actin filament organization and 9 coding genes in module 3 were actin-binding proteins; while 11 genes coding in the module 2 of WHHL model participated in myosin assembly. Totally there were 42 coding genes. 11 (14) up-regulated (down-regulated) in both models, 11 specifically up-regulated in cholesterol-fed NZW, and 6 specifically up-regulated in WHHL. These 42 coding genes were applied to STRING database to get an experimental or text-mining based gene-gene interaction network in which 18 of them showed tight interactions with high confidence 0.7. Twenty-three lncRNAs were added into the network based on the expression correlations between coding genes and lncRNAs (Fig. 5E, Supplemental Table 3). The actin-binding genes were downregulated in the cholesterol-fed NZW model, and their correlated lncRNAs also showed significant expression changes. However, most of the actin-binding related genes and lncRNAs did not have significant expression changes in WHHL. LncRNAs that involved in actin-binding protein regulation might be highly active in cholesterol-fed NZW models but not in WHHL models.

3. Discussion

We defined a relatively stringent set of 3736 multi-exonic long noncoding transcripts and 89,820 coding transcripts in rabbits. Those lncRNAs share the similar structure characteristics with human lncRNAs, such as shorter sequence length, longer exon and intron length, lower exon number, lower isoform number, relatively lower expression, higher tissue specificity compared to coding genes. The evolutionary conservation analysis also showed the similar conservation pattern with the previous observation in 17 species. Since rabbit is a widely used model organism in atherosclerosis, the catalog of rabbit lncRNAs would be extremely useful for studying the roles of lncRNAs in atherosclerosis. Since the RNA-seq library was constructed using a poly-A tailed method, our identification results were lack of non-polyadenylated lncRNAs. It will be enhanced as more complete transcriptomes are sequenced further.

Several experimental studies have shown evidences of lncRNA functions in atherosclerosis. Our conservation analysis with functionally known lncRNAs and association analysis with coding genes predicted the potential roles of rabbit lncRANs in many biological processes like lipid metabolism, immune response and muscle contraction. Furthermore, functional module recognition analysis of differential expressed lncRNAs and coding genes in two atherosclerosis models also revealed wildly participation of lncRNAs in inflammatory response and skeletal muscle system in aorta. Protein coding genes and lncRNAs in the most similar modules of two rabbit models were involved in immune responses. Besides, through comparing those less similar functional modules in two models, we found their functions majorly concentrated on skeletal muscle system while were diverse in different parts. LncRNAs in WHHL are myosin assembly related while actin assembly and binding protein related in NZW. In particular, the majority of actin binding protein associated lncRNAs are only differential expressed in NZW which indicate their potential special functions in regulation. Combining the overlapped or divergent lncRNA-gene pairs, four lncRNAs and their corresponding genes are differentially expressed: XLOC_000409 (gene: UBASH3B; divergent), XLOC_003013 (gene: PRR15; divergent), XLOC_006315 (gene: INSIG1; divergent) and XLOC_022497 (gene: FRS2; overlapped) which were strong disease related candidates. Collectively, the differential analysis suggests numerous roles of lncRNAs in atherosclerosis. Guilt-by-association analyses can be a good and effective method to study lncRNA functions compared to position associated method. But there is still space ahead because the annotations of rabbit genes are not comprehensive enough.

In summary, we have provided the first large-scale catalog of rabbit lncRNAs and provided a set of atherosclerosis-associated lncRNAs for further experimental validations.

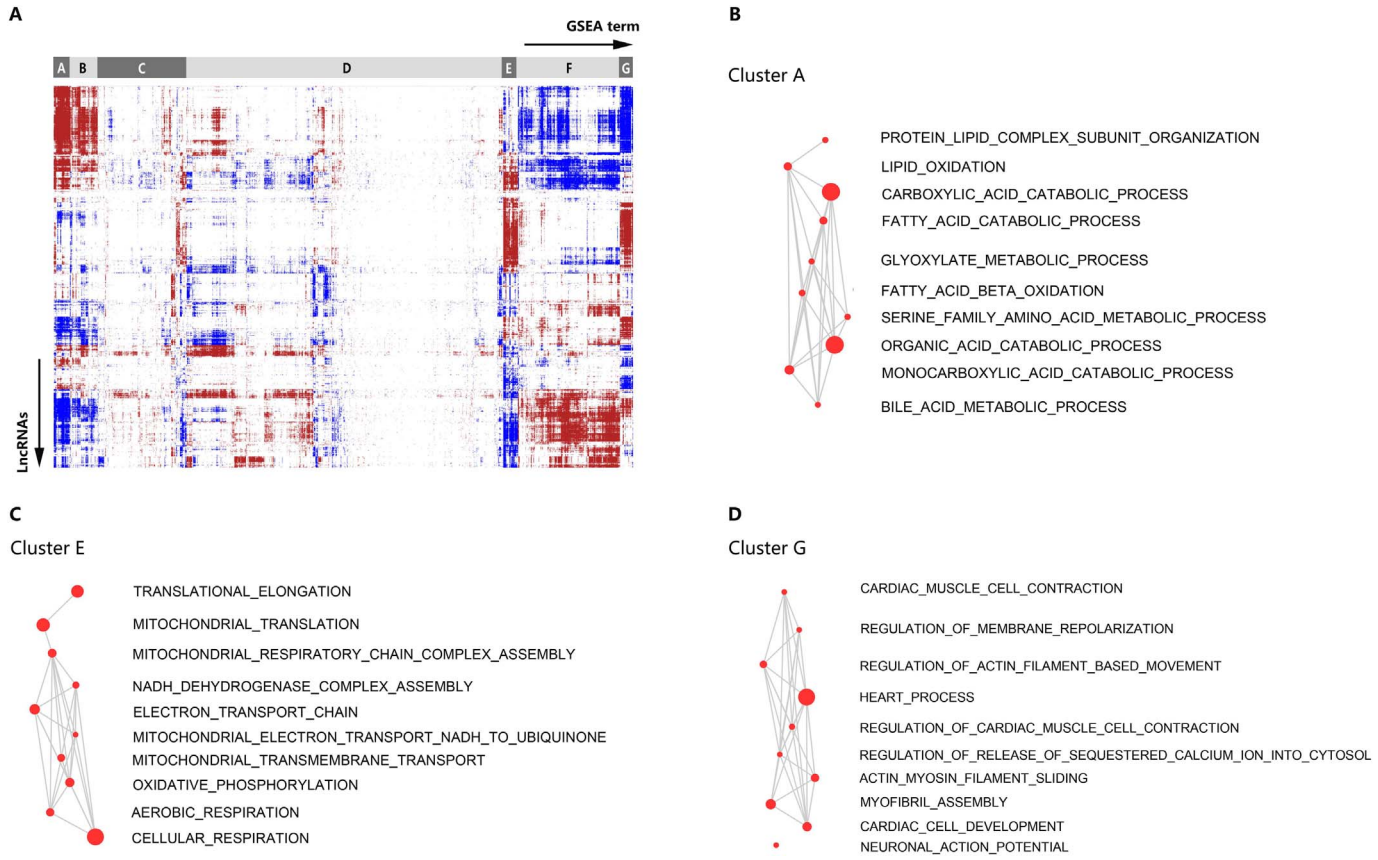

Fig. 4. Association matrix between lncRNAs and functional gene sets. (A) Expression based matrix of lncRNAs (column) and functional gene sets (row) by GSEA. Correlations are shown as positive (red), negative (blue) or no correlation (white). (B), (C), (D) highlight three dominant clusters associated with functions in lipid metabolism (B), immune response (C) and heart muscle contraction (D). GO BP terms in each cluster were visualized by a network. Each node is a GO term. Node size represents the number of genes in this GO term. Edge indicates that two GO terms share genes.

4. Methods

4.1. Transcriptome assembly

The RNA sequencing dataset came from the previous published paper [9] including 69 samples from four kinds of rabbits (wide-type NZW, cholesterol-fed NZW, wide-type JW, and LDLR-deficient WHHL), covering six tissues (aorta, coronary, heart, liver, kidney and embryo). RNA-seq derived reads were filtered by NGS QC Toolkit (v2.3.2) for quality control firstly. Reads were filtered once the quality score of > 75% bases was < 20. Bases in the 3′ end were trimmed if the quality were < 20, and reads < 40 bases after trimming were removed. The remaining high quality reads were aligned to rabbit reference genome (Ensemble OryCun2) by TopHat2 (version2.0.8). Then Cuffflinks (version2.0.2) was used to assemble transcripts and estimate expression values independently.

4.2. LncRNA classification and expression quantification

Once the transcriptome of each sample was assembled, we used Cuffmerge to merge all the transcripts. A class code was assigned to each Cufflinks transcript based on the type of match with reference transcript. Transcripts with class code "=" or "j" were regarded as potential coding transcripts, and transcripts with class code "i" or "u" were used to identify lncRNAs. Based on the intersections between lncRNAs and coding genes, we have used FEELnc [24] to classify rabbit lncRNAs into two types ("genic" and "intergenic"). A 10 kb window around each lncRNA was used to check for possible overlap with the nearest coding genes. When the lncRNA overlaps a coding gene, it is called "genic" (overlapped). Otherwise it is called "intergenic". Subsets of "intergenic" lncRNAs are called "directionally divergent" if the lncRNAs are transcribed in head to head orientation with gene partner. Expressions of transcripts and genes were quantified by the Fragments per Kilobase Million (FPKM). FPKM values were log2-transformed to get approximate normal distribution.

4.3. Coding potential assessment

PhyloCSF and CPC were used to assess coding potential for all rabbit transcripts. CPC utilized several biologically meaningful sequence features to build a support vector machine to predict protein-coding potential. Transcripts scored < 0 were regarded as noncoding. PhyloCSF analyzed a multispecies nucleotide sequence alignment to determine whether the transcript is protein-coding. Reference genomes of human and 6 rodent species were downloaded from UCSC database: human (hg38), mouse (mm10), rat (rn6), pika (ochPri3), squirrel (speTri2), guinea pig (cavPor3) and kangaroo rat (dipOrd1). Blastn was used to align rabbit transcripts with the genomes of another 7 species. The alignment results were input into PhyloCSF and assigned a score. PhyloCSF score less than "20" was set to define as noncoding. Taken together, transcripts were regarded as noncoding if they meet the standard of both softwares. When transcripts do not have homologous sequence in other species, PhyloCSF can't assess their coding potential and CPC predicted noncoding transcripts were also retained in the final results.

4.4. Tissue specificity

Tissue specificity index was calculated based on the following formula:

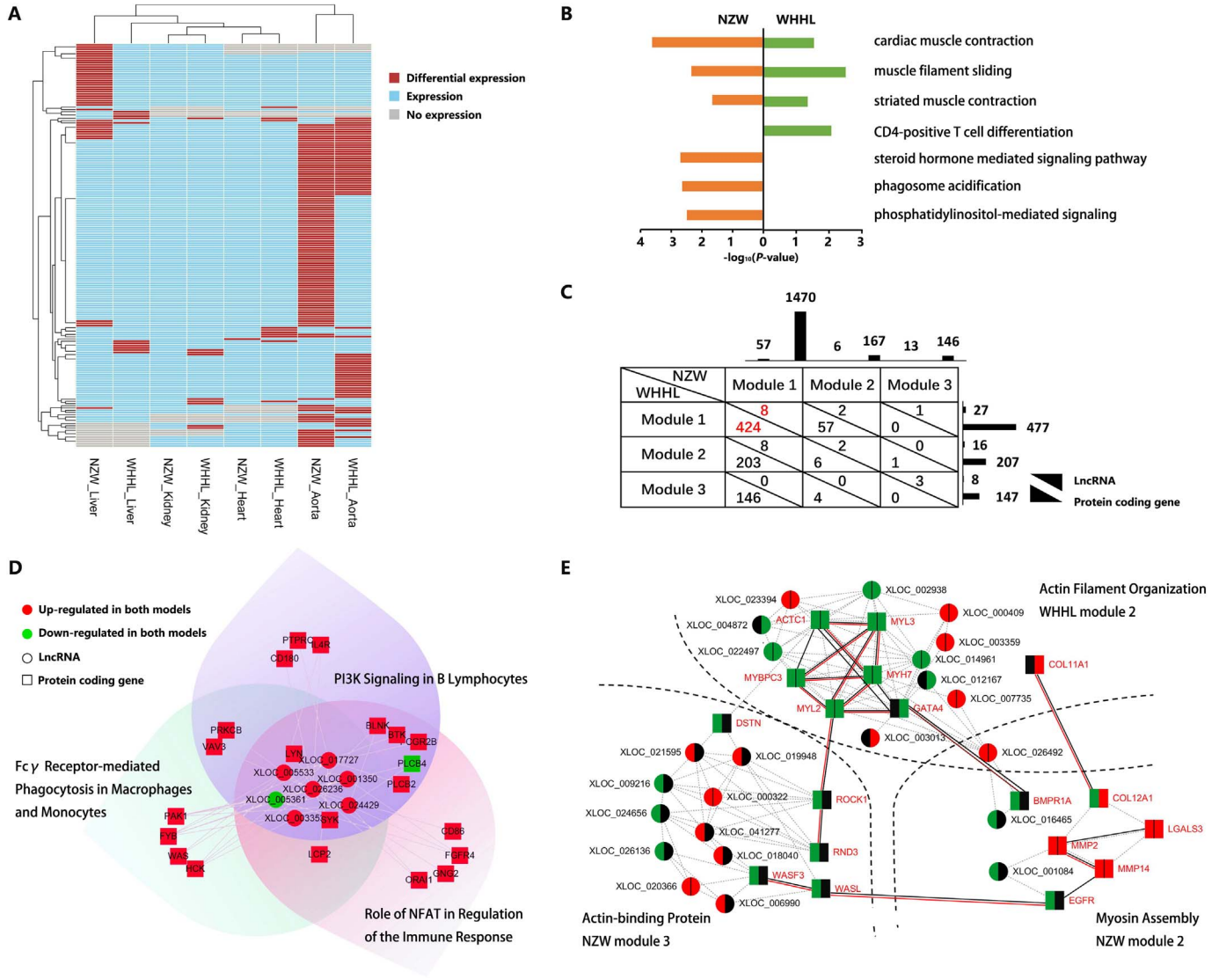

Fig. 5. LncRNA involved functional processes in atherosclerosis. (A) Heatmap of differential expressed lncRNAs in two kinds of atherosclerosis rabbit models (cholesterol-fed NZW, WHHL) in aorta, kidney, heart and liver. (B) Enriched GO categories of differentially expressed lncRNA in aorta in two models. (C) Comparison of lncRNAs and coding genes per functional module in two atherosclerosis models. Modules were generated from aorta differentially expressed genes by MCL method. Rows and columns show modules of WHHL and cholesterol-fed NZW models, respectively. The black bars represent the number of lncRNAs (left) and coding genes (right) in each module. Numbers in cell indicate shared lncRNAs (upper right) and coding genes (lower left) of each MCL module pair in two models. Cell with red number represents the most similar pair of MCL module. (D) Network of shared lncRNAs and coding genes in the most similar module pair in aorta. Edges are expression correlated lncRNAs and coding genes. (E) Network of skeletal muscle system in aorta. The left and right semicircle of each node shows expression change of cholesterol-fed NZW and WHHL models compared to the corresponding normal controls. Edges between coding gene and lncRNA represent expression correlations (grey dash line). Edges between gene and gene represent protein-protein interactions, which were supported by experimental evidence or txt-mining (black line) and experimental evidence only (red line).

$$\text{Index} = \frac{\sum_{i=1}^{n}\left(1 - \frac{exp_i}{exp_{max}}\right)}{n - 1.}$$

where n is the number of tissue (here we regarded tissue in different breeds as different tissues [2]); exp_i stands for the mean expression value of tissue i; exp_{max} is the maximum expression value across all tissues.

To better present the expression pattern of genes using heatmap, expression value was scaled across all 69 samples:

$$\text{E} = \frac{log_2(1 + FPKM_i)}{\sum_{i=1}^{69} log_2(1 + FPKM_i)}$$

where $FPKM_i$ stands for the original FPKM value for sample i. A pseudo-number 1 was added in order to obtain non negative expression value.

4.5. Functional annotation by GSEA

Spearman correlation coefficients (SCC) were calculated between all lncRNA and protein-coding genes. For each lncRNA, protein-coding gene was ranked based on the correlation coefficients. This pre-ranked gene list was input into Gene Set Enrichment Analysis (GSEA) software. GSEA returned the Enrichment Score (ES) and false discovery rate (FDR) for each Gene Ontology Biological Progress (GO BP). An ES matrix was conducted, whose rows correspond to lncRNA and columns correspond to the significantly enriched GO BP terms (FDR < 0.05). Hierarchical cluster analysis was used to cluster functionally similar lncRNA into groups.

4.6. Differential expression and functional module

Raw counts were normalized by DESeq package (version1.24.0),

and differential expressed lncRNAs and coding genes were detected by Binomial Test (FDR < 0.1 and fold change > 2). The co-expression network was built, whose nodes are differential expressed lncRNAs or coding genes, edges are correlated lncRNA-lncRNA, gene-gene, and lncRNA-gene. Markov Cluster Algorithm (MCL) was applied to the co-expression network to recognize potential functional modules. In order to determine the optimal modules, we used different correlation cutoff to build co-expression networks and run MCL. The cutoff of SCC started from 0.2 and increase 0.05 each time. The optimal cutoff was chosen to make the number of modules larger than 10 and the percentage of singleton and excluded lncRNA number as lower as possible (Supplemental Fig. 5).

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbadis.2017.12.040.

Transparency document

The http://dx.doi.org/10.1016/j.bbadis.2017.12.040 associated with this article can be found, in online version.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (31501077, 31771472), National Grand Program on Key Infectious Diseases (2015ZX10004801-005), National High-Tech R&D Program (863) 2015AA020105.

Author contributions

H.L., Yixue L. and J.L. designed the study. J.L., Qianlan Y. and S.H. performed the lncRNA identification. J.L., Liguang Y. and P.L. performed the function analysis. Fangyoumin F. and Chuhua Y. performed the IPA analysis. H.L. and J.L. wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

References

[1] T. Derrien, R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D.G. Knowles, The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression, Genome Res. 22 (2012) 1775–1789.

[2] M.N. Cabili, C. Trapnell, L. Goff, M. Koziol, B. Tazon-Vega, A. Regev, J.L. Rinn, Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses, Genes Dev. 25 (2011) 1915–1927.

[3] H. Hezroni, D. Koppstein, M.G. Schwartz, A. Avrutin, D.P. Bartel, I. Ulitsky, Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species, Cell Rep. 11 (2015) 1110–1122.

[4] E.J. Benjamin, M.J. Blaha, S.E. Chiuve, M. Cushman, S.R. Das, R. Deo, S.D. de Ferranti, J. Floyd, M. Fornage, C. Gillespie, Heart disease and stroke statistics—2017 update: a report from the American Heart Association, Circulation 135 (2017) e146–e603.

[5] D. Steinberg, In celebration of the 100th anniversary of the lipid hypothesis of atherosclerosis, J. Lipid Res. 54 (2013) 2946–2949.

[6] C.S. Robbins, I. Hilgendorf, G.F. Weber, I. Theurl, Y. Iwamoto, J.L. Figueiredo, R. Gorbatov, G.K. Sukhova, L.M. Gerhardt, D. Smyth, C.C. Zavitz, E.A. Shikatani, M. Parsons, N. van Rooijen, H.Y. Lin, M. Husain, P. Libby, M. Nahrendorf, R. Weissleder, F.K. Swirski, Local proliferation dominates lesional macrophage accumulation in atherosclerosis, Nat. Med. 19 (2013) 1166–1172.

[7] J. Fan, S. Kitajima, T. Watanabe, J. Xu, J. Zhang, E. Liu, Y.E. Chen, Rabbit models for the study of human atherosclerosis: from pathophysiological mechanisms to translational medicine, Pharmacol. Ther. 146 (2015) 104–119.

[8] F.R. Kapourchali, L.C. Gangadaran Surendiran, E. Uitz, B. Bahadori, M.H. Moghadasian, Animal models of atherosclerosis, World J. Clin. Cases: WJCC 2 (2014) 126.

[9] Z. Wang, J. Zhang, H. Li, J. Li, M. Niimi, G. Ding, H. Chen, J. Xu, H. Zhang, Z. Xu, Hyperlipidemia-associated gene variations and expression patterns revealed by whole-genome and transcriptome sequencing of rabbit models, Sci. Rep. 6 (2016) 26942.

[10] G. Wu, J. Cai, Y. Han, J. Chen, Z.-P. Huang, C. Chen, Y. Cai, H. Huang, Y. Yang, Y. Liu, LincRNA-p21 regulates neointima formation, vascular smooth muscle cell proliferation, apoptosis and atherosclerosis by enhancing p53 activity, Circulation (2014) 114.011675 (CIRCULATIONAHA).

[11] C. Rodríguez-Mateo, B. Torres, G. Gutiérrez, J.A. Pintor-Toro, Downregulation of Lnc-Spry1 mediates TGF-β-induced epithelial–mesenchymal transition by transcriptional and posttranscriptional regulatory mechanisms, Cell Death Differ. 24 (2017) 785–797.

[12] C.E. Burd, W.R. Jeck, Y. Liu, H.K. Sanoff, Z. Wang, N.E. Sharpless, Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk, PLoS Genet. 6 (2010) e1001233.

[13] L. Kong, Y. Zhang, Z.-Q. Ye, X.-Q. Liu, S.-Q. Zhao, L. Wei, G. Gao, CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine, Nucleic Acids Res. 35 (2007) W345–W349.

[14] M.F. Lin, I. Jungreis, M. Kellis, PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions, Bioinformatics 27 (2011) i275–i282.

[15] R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, Pfam: the protein families database, Nucleic Acids Res. 42 (2013) D222–D230.

[16] A. Pauli, E. Valen, M.F. Lin, M. Garber, N.L. Vastenhouw, J.Z. Levin, L. Fan, A. Sandelin, J.L. Rinn, A. Regev, Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis, Genome Res. 22 (2012) 577–591.

[17] M.K. Iyer, Y.S. Niknafs, R. Malik, U. Singhal, A. Sahu, Y. Hosono, T.R. Barrette, J.R. Prensner, J.R. Evans, S. Zhao, The landscape of long noncoding RNAs in the human transcriptome, Nat. Genet. 47 (2015) 199–208.

[18] Y. Zhao, H. Li, S. Fang, Y. Kang, Y. Hao, Z. Li, D. Bu, N. Sun, M.Q. Zhang, R. Chen, NONCODE 2016: an informative and valuable data source of long non-coding RNAs, Nucleic Acids Res. 44 (2016) D203–D208.

[19] P.P. Amaral, M.B. Clark, D.K. Gascoigne, M.E. Dinger, J.S. Mattick, lncRNAdb: a reference database for long noncoding RNAs, Nucleic Acids Res. 39 (2010) D146–D151.

[20] I.A. Mitchell Guttman, M. Garber, C. French, M.F. Lin, D. Feldser, M. Huarte, O. Zuk, B.W. Carey, J.P. Cassady, M.N. Cabili, Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals, Nature 458 (2009) 223.

[21] J. Hauke, T. Kossowski, Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data, Quaest. Geogr. 30 (2011) 87.

[22] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. 102 (2005) 15545–15550.

[23] K.J. Moore, I. Tabas, Macrophages in the pathogenesis of atherosclerosis, Cell 145 (2011) 341–355.

[24] V. Wucher, F. Legeai, B. Hédan, G. Rizk, L. Lagoutte, T. Leeb, V. Jagannathan, E. Cadieu, A. David, H. Lohi, FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome, Nucleic Acids Res. 45 (2017) e57.