SCIENCE CHINA
Life Sciences

THEMATIC ISSUE: Vascular homeostasis and injury-reconstruction
• RESEARCH PAPER •

August 2014 Vol.57 No.8: 852–857
doi: 10.1007/s11427-014-4692-4

A bioinformatics method for predicting long noncoding RNAs associated with vascular disease

LI JianWei[1,2†*], GAO Cheng[3,4†], WANG YuChen[1,3,4], MA Wei[2,3,4], TU Jian[2,3,4], WANG JunPei[2,3,4], CHEN ZhenZhen[2,3,4], KONG Wei[3,4] & CUI QingHua[2,4*]

[1]*Laboratory of Translational Biomedicine Informatics, School of Computer Science, Hebei University of Technology, Tianjin 300401, China;*
[2]*Department of Biomedical Informatics, School of Basic Medical Sciences, Peking University, Beijing 100191, China;*
[3]*Department of Physiology and Pathophysiology, School of Basic Medical Sciences, Peking University, Beijing 100191, China;*
[4]*MOE Key Laboratory of Cardiovascular Sciences, Peking University, Beijing 100191, China*

Received April 30, 2014; accepted May 20, 2014

Long noncoding RNAs (lncRNAs) play important roles in human diseases including vascular disease. Given the large number of lncRNAs, however, whether the majority of them are associated with vascular disease remains unknown. For this purpose, here we present a genomic location based bioinformatics method to predict the lncRNAs associated with vascular disease. We applied the presented method to globally screen the human lncRNAs potentially involved in vascular disease. As a result, we predicted 3043 putative vascular disease associated lncRNAs. To test the accuracy of the method, we selected 10 lncRNAs predicted to be implicated in proliferation and migration of vascular smooth muscle cells (VSMCs) for further experimental validation. The results confirmed that eight of the 10 lncRNAs (80%) are validated. This result suggests that the presented method has a reliable prediction performance. Finally, the presented bioinformatics method and the predicted vascular disease associated lncRNAs together may provide helps for not only better understanding of the roles of lncRNAs in vascular disease but also the identification of novel molecules for the diagnosis and therapy of vascular disease.

vascular disease, lncRNAs, bioinformatics

Citation: Li JW, Gao C, Wang YC, Ma W, Tu J, Wang JP, Chen ZZ, Kong W, Cui QH. A bioinformatics method for predicting long noncoding RNAs associated with vascular disease. Sci China Life Sci, 2014, 57: 852−857, doi: 10.1007/s11427-014-4692-4

In recent years, a surprising finding in the analysis of human transcriptome is that mRNAs only account for a small portion of the genome transcripts [1]. The majority of the human genome transcripts are noncoding RNAs, especially, long noncoding RNAs (lncRNAs) [2]. Historically, people often argue against the functionality of lncRNAs [3] because they normally tend to show low cross-species conservation, low expression levels and high tissue specificity. Recently, however, increasing evidence has suggested that a number of lncRNAs have important and diverse functions [4]. Therefore, it will be no surprise the dysfunction of lncRNAs is associated with a wide spectrum of disease, including cardiovascular disease [5] and cancer [6]. According to the statistics of the long noncoding RNA disease database (LncRNADisease, http://www.cuilab.cn/lncrnadisease) [7], more than 200 diseases were reported to be associated with lncRNAs and more than 250 lncRNAs were reported to have roles in disease. Therefore, it is believed that lncRNAs are becoming a large class of novel molecules for disease diagnosis and therapy [8]. Vascular disease represents one class of serious disease that causes lethal death in the world [9]. Given the large number of lncRNAs, however, the relationship between the majority of lncRNAs and

†Contributed equally to this work
*Corresponding author (email: cuiqinghua@bjmu.edu.cn; kjl205@163.com)

human disease including vascular disease remains unknown. Therefore, it is increasingly emergent and important to identify the roles of lncRNAs in vascular disease by developing bioinformatics methods.

For the bioinformatics methods to predict novel lncRNA-disease associations, we previously presented a genomic location based bioinformatics method [7]. Moreover, two bioinformatics methods based on co-expression of lncRNA and protein-coding genes were presented for cancer [10] and human disease [11], respectively. In addition, a method of Laplacian Regularized Least Squares for lncRNA-disease association (LRLSLDA) based on data from the LncRNADisease database and lncRNA expression profile was developed [12]. More recently, a network based analysis of lncRNA-disease association data in the LncRNADisease database revealed a number of regular patterns between lncRNAs and disease [13], suggesting that it is feasible to predict novel lncRNA-disease associations. Indeed, in the same study, the authors developed a network-based method to predict novel lncRNA-disease associations [13]. These studies and the presented bioinformatics methods together provide great helps in understanding the roles of lncRNAs in human disease and in finding novel lncRNA-disease associations. Of course, limitations exist in the above studies. For example, the genomic location based method used a genomic distance of 2 kb as the cutoff to finding the neighbor genes of an lncRNA. Moreover, the original method only focused on disease-associated genes but neglected disease-associated gene ontology (GO) and gene mutation (e.g., SNPs). Comprehensive analysis of lncRNA SNPs has showed that SNPs in lncRNAs could contribute to disease susceptibility [14]. For the co-expression based methods, only a small number of lncRNAs have matched tissues with protein-coding genes and lncRNAs does not always have similar function with their co-expressed protein-coding genes. In addition, the lncRNA-disease network methods only focus on a limited number of lncRNAs (~260) contained in the LncRNADisease database and cannot be applied to the majority of human lncRNAs. Therefore, more methods are emergently needed. Here, we focus on vascular disease and improved the genomic location based method in two aspects. One is that we extended the genomic distance to 50 kb, which is suggested to define miRNA clusters in the field of miRNAs, another class of noncoding RNAs. It is known that the miRNAs in one cluster often have similar expression, functions [15], and disease [16]. That is, we obtained vascular function associated genes first and then identified the lncRNAs within the regions of 50 kb from any of the vascular disease genes. In addition, we considered the vascular related GO terms and vascular disease related SNPs. Together, these lncRNAs are considered to be vascular disease associated lncRNAs. Finally, to evaluate the accuracy of the predictions, we randomly select 10 lncRNAs predicted to be associated with VSMC proliferation and migration for further biological experiment validation. As a result, we confirmed that the prediction has a high accuracy.

1 Materials and methods

1.1 The genomic data used in this study

We downloaded the genomic location data of 32108 human lncRNAs from the LNCipedia database [17] (http://www.lncipedia.org/). We downloaded the data of gene-disease from the GAD database [18] (http://geneticassociationdb.nih. gov/). We downloaded the GO from NCBI and disease SNPs from the NHGRI GWAS Catalog [19] (http://www.genome.gov/gwastudies/). We downloaded the genomic location data of reference genes from UCSC [20] (http://genome.ucsc.edu/).

1.2 The flowchart to predict vascular disease associated lncRNAs

As shown in Figure 1, we manually extracted vascular disease associated genes from the GAD dataset and vascular disease associated SNPs from the NHGRI GWAS Catalog dataset. We then used an in-house java program to extract vascular related GO terms using expert knowledge combined with the following keywords, angiogenesis, angiostatin, arterial, artery, blood, circulation, vascular, vasculogenesis, vasoactive, vasoconstriction, vasodilation, vasomotion, vasopressin, VEGF, vein, and vessel. We next identified the lncRNAs that are within 50 kb from genes associated with vascular disease related genes and genes associated vascular related GO terms. We also identified the lncRNAs that host vascular related SNPs. Together we considered these lncRNAs as putative vascular disease associated lncRNAs.

1.3 The model of VSMC proliferation and migration

VSMCs are highly differentiated cells, but they can undergo phenotypic switch from contractile to dedifferentiated phenotype such as synthetic, inflammatory and osteo/

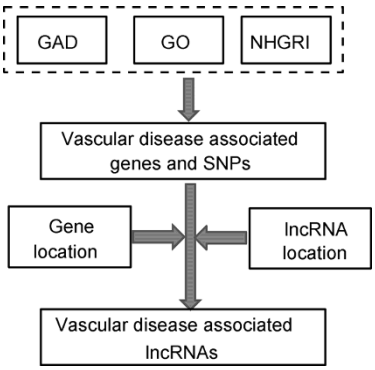

Figure 1 Flowchart of the genomic location based bioinformatics method for the prediction of novel lncRNA-vascular disease associations.

chondrocytic phenotype in response to the change of extracellular matrix. VSMCs' phenotypic transition plays a critical role in angiogenesis and restenosis [21]. During this process, various growth factors and cytokines such as platelet-derived growth factor-BB (PDGF-BB), interleukin-1, tumor necrosis factor α (TNF-α) and fibroblast growth factor (FGF) are markedly increased [22]. PDGF-BB and its signaling is a key regulator to mediate VSMC proliferation and migration. Inflammatory cytokine TNF-α is powerful to promote the secretion of multiple chemokines of VSMCs and regulate VSMC proliferation and migration in aortic wall injury site. Transcriptional activation of VSMCs marker genes was markedly suppressed by using PDGF-BB and TNF-α treatment with primary VSMCs [22,23].

Recombinant tumor necrosis factors (TNF-α and PDGF-BB) were purchased from Peprotech Ltd. (Rocky Hill, USA). VSMCs were isolated from the thoracic aortic arteries of Sprague-Dawley rats (body weight 150–180 g) as described previously [24] and cells at passages 4–8 were used in all experiments.

1.4 Quantitative real-time PCR analysis

To test the accuracy of the presented method, here we selected 10 lncRNAs predicted to be associated with VSMC proliferation and migration for further experiment validation. We used quantitative real-time PCR analysis to evaluate whether the predicted lncRNAs are deregulated in the treated VSMCs. Real-time PCR amplification involved use of an Mx3000 Multiplex Quantitative PCR System (Stratagene Corp, USA) and Eva Green reagent normalized to that of the internal control β-actin. The 10 candidate lncRNAs and their primer sequences of target genes are listed in Table 1. All amplification reactions were carried out over 40 cycles (an initial stage of 7 min at 94°C, then a three-step program of 30 s at 94°C, 30 s at 60°C and 30 s at 72°C) and were performed in duplicate.

Table 1 The 10 candidate lncRNAs and their primers for qRT-PCR validation experiments

lnc RNA name	Sense and anti-sense primers	Primer sequence
lnc-AC016251	Sence	5′-AATCTCTGGCCTTCGTG-3′
	Antisence	5′-CTTCGGATCTTCGTGTAGCTG-3′
lnc-AGPAT2	Sence	5′-CAGCTTTGCCCTATCC-3′
	Antisence	5′-TGATGAGGGTTCTCTGCGTCT-3′
lnc-AK1	Sence	5′-ACAGAACCTGTCATCGCCTTC-3′
	Antisence	5′-CGGCTCCAGCGTTGCTACTTT-3′
lnc-ATP6V1E2	Sence	5′-CTTTGCGTTCTGTGAGTGTGC-3′
	Antisence	5′-GGCGTTCTCTGGAGTATTGGA-3′
lnc-CLORF168	Sence	5′-CACTCTACCTCGCTGGC-3′
	Antisence	5′-CTGACATTCCATTGGCTAAAG-3′
lnc-CDK9	Sence	5′-AGGAAGAGAGGCGAATAGCGT-3′
	Antisence	5′-TCCCACCTCCGCTGAGTCGT-3′
lnc-EGFL7	Sence	5′-GATGGTGGGAAGCGTTCAGAC-3′
	Antisence	5′-CGCCTCCAGGACACACTTACT-3′
lnc-FPGS	Sence	5′-TAACTGGAGCAGGAACTCG-3′
	Antisence	5′-AGTAGCTGGGACTATGGGTGT-3′
lnc-RFPL4B	Sence	5′-TCCGTTGTGCCTTTAGAAC-3′
	Antisence	5′-TACTCAGCGAACACGTACACT-3′
lnc-TGFBR2	Sence	5′-TAATCATCCTAGAAGCCCTAC-3′
	Antisence	5′-GCCGACCTTGGGTGATACAC-3′

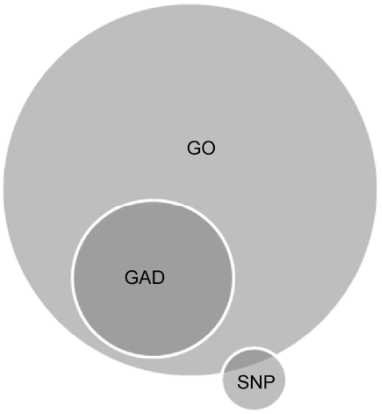

Figure 2 A pie chart for the distribution of predicted vascular disease associated lncRNAs based on the three types (GO-based, GAD-based, and SNP-based) of information.

2 Results

2.1 Putative vascular disease associated lncRNAs

The predicted total vascular disease associated lncRNAs, their sequence, the associated disease, GO terms, and SNPs are available at the download page of the LncRNADisease database (http://www.cuilab.cn/lncrnadisease). There are 2983 lncRNAs, 107 lncRNAs, and 72 lncRNAs in the GAD-based predictions, the GO-based predictions, and the SNP-based predictions, respectively. As shown in Figure 2, all of the 107 GAD-based predictions are included in the GO-based predictions, whereas only 12 of the SNP-based predictions are included in the GO-based predictions. There is no overlap between the GAD-based predictions and the SNP-based predictions.

It will be interesting to find whether some known vascular disease associated lncRNAs are successfully predicted by our method. For this purpose, we first curated the known vascular disease associated lncRNAs reported in previous studies [25–27]. They are Lnc-Ang362, eNOS-AS, TIE-AS, MALAT1, and ANRIL. We then matched these lncRNAs with the lncRNAs predicted to be associated with vascular disease. As a result, three (15 antisense transcripts close to eNOS, lnc-CDK5-1:1, lnc-CDK5-2:1–lnc-CDK5-2:14; three antisense transcripts close to TIE, lnc-ELOVL1-1:1, lnc-ELOVL1-1:2, and lnc-ELOVL1-1:3; 16 transcripts of ANRIL, lnc-MTAP-1:1–lnc-MTAP-1:16) of the five lncRNAs were found, suggesting that these lncRNAs were

successfully predicted. For the two lncRNAs that were not predicted, Lnc-Ang362 was not included in the lncRNA sequence database (LNCipedia) used in this study. For MALAT1, its neighboring gene, *SCYL1*, is not a known vascular-related gene based on GAD and GO datasets. Therefore, these two genes were not predicted to be associated with vascular diseases by this method. This result suggests that the presented method has reliable prediction accuracy.

2.2 Validation of putative lncRNAs associated VSMCs proliferation and migration

To evaluate the accuracy of the predictions, we randomly selected 10 lncRNAs associated with VSMCs proliferation and migration for further validation by biological experiments. For this purpose, here we built two cell models, VSMCs treated by PDGF-BB and VSMCs treated by TNF-α. We next used qRT-PCR to evaluate whether the candidate lncRNAs are deregulated in the treated cells compared with the untreated ones. As a result, we found that 70% (7/10) of the candidate lncRNAs showed significant deregulation (P<0.05) in the PDGF-BB treated VSMCs (Figure 3). Moreover, 50% (5/10) of the candidate lncRNAs showed significant deregulation (P<0.05) in the TNF-α treated VSMCs (Figure 4). Together 80% of the candidate lncRNAs are significantly deregulated. The results suggest that these putative lncRNAs could be indeed involved in VSMCs proliferation and migration, which further suggests that the presented method has a reliable accuracy. Moreover, by GO-based method, we know that all of the eight lncRNAs are putatively involved in blood vessel remodeling, blood vessel morphogenesis, artery morphogenesis, and blood vessel development.

3 Discussion

lncRNAs represent one large class of important noncoding molecules, which play critical roles in human diseases. Vascular disease is one class of serious disease that leads to lethal death in the world. Therefore, it is increasingly emergent and important to rapidly identify lncRNAs that are implicated in vascular disease. In this study, we presented a simple genomic location based *in-silico* method to predict novel associations of lncRNAs and vascular disease. An independent biological experiment confirmed that the presented method has a reliable accuracy. Of course, limitations exist in the current method. For example, not all of the lncRNAs have at least one neighbor gene within 50 kb distance. Moreover, lncRNAs are not always functional-

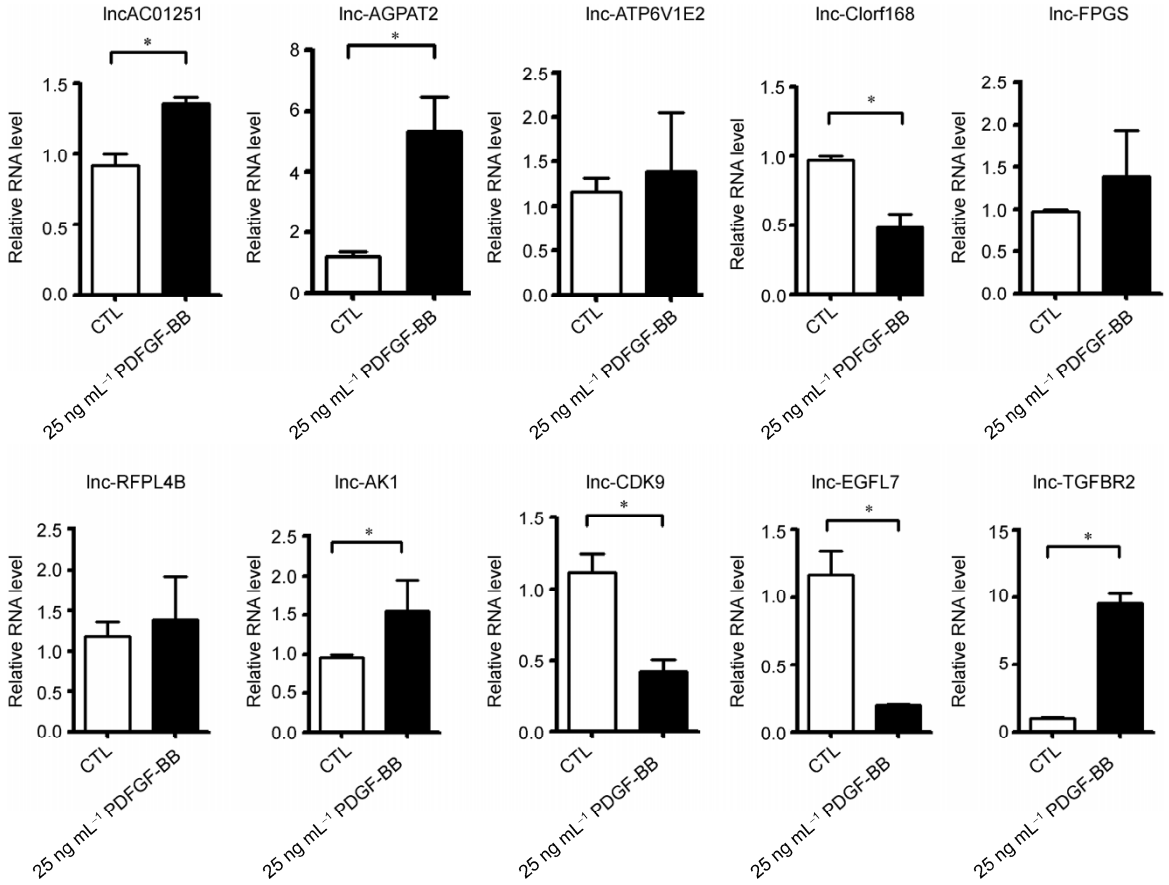

Figure 3 Validation results of the 10 candidate lncRNAs associated with VSMC proliferation and migration in the PDGF-BB treated VSMCs. * represents significance (P<0.05).

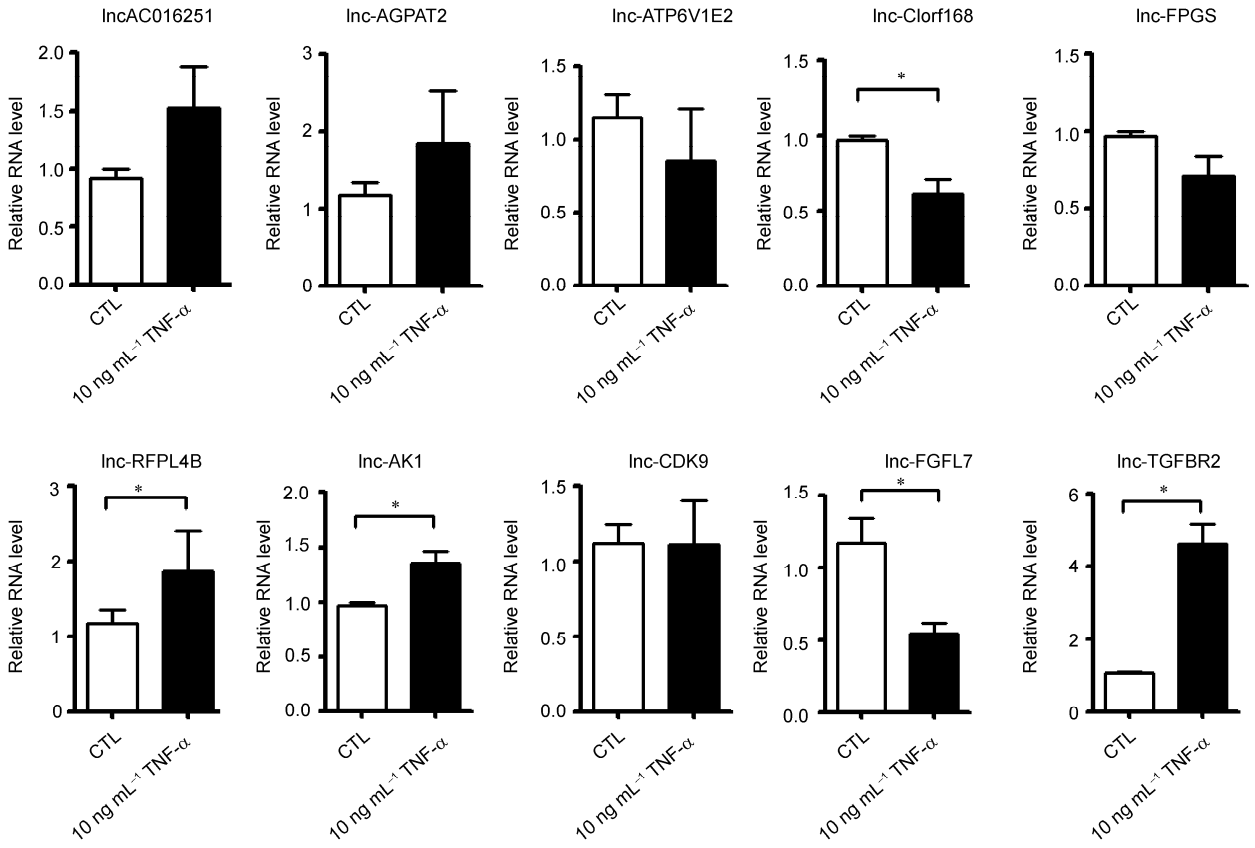

Figure 4 Validation results of the 10 candidate lncRNAs associated with VSMC proliferation and migration in the TNF-α treated VSMCs. * represents significance (P<0.05).

ly related with their neighbor genes. In addition, no statistical tests were used in the current method. Thus, the method seems to be less quantitative. Although limitations exist, we believe that the presented method and the predicted vascular disease associated lncRNAs provide researchers valuable resources for not only better understanding the roles of lncRNAs in vascular disease but also identifying novel potential biomarkers and drug targets for vascular disease diagnosis and therapy.

This work was supported by the National Natural Science Foundation of China (91339106) and National High Technology Research and Development Program of China (2014AA021102).

1 Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, Gerstein M, Snyder M. Global identification of human transcribed sequences with genome tiling arrays. Science, 2004, 306: 2242–2246
2 Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, Bell I, Cheung E, Drenkow J, Dumais E, Patel S, Helt G, Ganesh M, Ghosh S, Piccolboni A, Sementchenko V, Tammana H, Gingeras TR. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science, 2007, 316: 1484–1488
3 Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell, 2009, 136: 629–641
4 Schonrock N, Harvey RP, Mattick JS. Long noncoding RNAs in cardiac development and pathophysiology. Circ Res, 2012, 111: 1349–1362
5 Klattenhoff CA, Scheuermann JC, Surface LE, Bradley RK, Fields PA, Steinhauser ML, Ding H, Butty VL, Torrey L, Haas S, Abo R, Tabebordbar M, Lee RT, Burge CB, Boyer LA. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell, 2013, 152: 570–583
6 Cheetham SW, Gruhl F, Mattick JS, Dinger ME. Long noncoding RNAs and the genetics of cancer. Br J Cancer, 2013, 108: 2419–2425
7 Chen G, Wang Z, Wang D, Qiu C, Liu M, Chen X, Zhang Q, Yan G, Cui Q. LncRNADisease: a database for long-non-coding RNA-associated diseases. Nucleic Acids Res, 2013, 41: D983–986
8 Wapinski O, Chang HY. Long noncoding RNAs and human disease. Trends in cell biology, 2011, 21: 354–361
9 Roberts A. Vascular disease: disparities in treatment and mortality from ruptured AAA between england and the USA. Nat Rev Cardiol, 2014, 11: 247
10 Sun L, Luo H, Liao Q, Bu D, Zhao G, Liu C, Liu Y, Zhao Y. Systematic study of human long intergenic non-coding RNAs and their impact on cancer. Sci China Life Sci, 2013, 56: 324–334
11 Liu MX, Chen X, Chen G, Cui QH, Yan GY. A computational framework to infer human disease-associated long noncoding RNAs. PLoS ONE, 2014, 9: e84408
12 Chen X, Yan GY. Novel human lncRNA-disease association inference based on lncRNA expression profiles. Bioinformatics, 2013, 29: 2617–2624
13 Yang X, Gao L, Guo X, Shi X, Wu H, Song F, Wang B. A network based method for analysis of lncRNA-disease associations and prediction of lncRNAs implicated in diseases. PLoS ONE, 2014, 9: e87797

14 Chen G, Qiu C, Zhang Q, Liu B, Cui Q. Genome-wide analysis of human SNPs at long intergenic noncoding RNAs. Hum Mutat, 2013, 34: 338–344
15 Wang D, Wang J, Lu M, Song F, Cui Q. Inferring the human microrna functional similarity and functional network based on microRNA-associated diseases. Bioinformatics, 2010, 26: 1644–1650
16 Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, Cui Q. An analysis of human microRNA and disease associations. PLoS ONE, 2008, 3: e3420
17 Volders PJ, Helsens K, Wang X, Menten B, Martens L, Gevaert K, Vandesompele J, Mestdagh P. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res, 2013, 41: D246–251
18 Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. Nat Genet, 2004, 36: 431–432
19 Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res, 2014, 42: D1001–1006
20 Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. The ucsc table browser data retrieval tool. Nucleic Acids Res, 2004, 32: D493–496
21 Alexander MR, Owens GK. Epigenetic control of smooth muscle cell differentiation and phenotypic switching in vascular development and disease. Ann Rev Physiol, 2012, 74: 13–40
22 Owens GK, Kumar MS, Wamhoff BR. Molecular regulation of vascular smooth muscle cell differentiation in development and disease. Physiol Rev, 2004, 84: 767–801
23 Wang L, Zheng J, Bai X, Liu B, Liu CJ, Xu Q, Zhu Y, Wang N, Kong W, Wang X. Adamts-7 mediates vascular smooth muscle cell migration and neointima formation in balloon-injured rat arteries. Circ Res, 2009, 104: 688–698
24 Ross R. The smooth muscle cell. II. Growth of smooth muscle in culture and formation of elastic fibers. J Cell Biol, 1971, 50: 172–186
25 Aryal B, Rotllan N, Fernandez-Hernando C. Noncoding RNAs and atherosclerosis. Curr Atheroscler Rep, 2014, 16: 407
26 Leung A, Trac C, Jin W, Lanting L, Akbany A, Saetrom P, Schones DE, Natarajan R. Novel long noncoding RNAs are regulated by angiotensin II in vascular smooth muscle cells. Circ Res, 2013, 113: 266–278
27 Michalik KM, You X, Manavski Y, Doddaballapur A, Zornig M, Braun T, John D, Ponomareva Y, Chen W, Uchida S, Boon RA, Dimmeler S. Long noncoding RNA MALAT1 regulates endothelial cell function and vessel growth. Circ Res, 2014, 114: 1389–1397