Gene 575 (2016) 303–312

Contents lists available at ScienceDirect

Gene

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

Research paper

Comparative transcriptome analysis of atrial septal defect identifies dysregulated genes during heart septum morphogenesis

Wenju Wang [1], Zhaoyi Niu [1], Yi Wang, Yaxiong Li, Honglin Zou, Li Yang, Mingyao Meng, Chuanyu Wei, Qinrui Li, Le Duan, Yanhua Xie, Yayong Zhang, Yu Cao, Shen Han, Zongliu Hou *, Lihong Jiang **

[a] Yan'an Affiliated Hospital of Kunming Medical University, Kunming 650051, Yunnan, People's Republic of China
[b] Kunming Yan'an Hospital, Kunming 650051, Yunnan, People's Republic of China
[c] Institute of Cardiovascular Surgery of Yunnan Province, Kunming 650051, Yunnan, People's Republic of China

ARTICLE INFO

Article history:
Received 23 July 2015
Received in revised form 28 August 2015
Accepted 2 September 2015
Available online 14 September 2015

Keywords:
Congenital heart disease
Atrial septal defect
Transcriptome
Deep sequencing
Bioinformatic analysis
Heart development

ABSTRACT

Congenital heart disease (CHD) is one of most common birth defects, causing fetal loss and death in newborn all over the world. Atrial and ventricular septal defects were the most common CHD subtypes in most districts. During the past decades, several genes were identified to control atrial septum formation, and mutations of these genes can cause cardiac septation defects. However, the pathogenic mechanism of ASD on transcriptional levels has not been well elucidated yet. Herein, we performed comparative transcriptome analysis between normal and atrial septal defect (ASD) patients by Illumina RNA sequencing (RNA-seq). Advanced bioinformatic analyses were employed to identify dysregulated genes in ASD. The results indicated that cardiac specific transcriptional factors (GATA4 and NKX2-5), extracellular signal molecules (VEGFA and BMP10) and cardiac sarcomeric proteins (MYL2, MYL3, MYH7, TNNT1 and TNNT3) were downregulated in ASD which may affect heart atrial septum formation, cardiomyocyte proliferation and cardiac muscle development. Importantly, cell cycle was dominant pathway among downregulated genes, and decreased expression of the proteins included in cell cycle may disturb cardiomyocyte growth and differentiation during atrial septum formation. Our study provided evidences of understanding pathogenic mechanism of ASD and resource for validation of CHD genomic studies.

1. Introduction

Congenital heart disease (CHD) is one of most common birth defects, causing fetal loss and death in newborns all over the world. CHD accounts for 28% of all major congenital anomalies, affecting about 0.8% of live birth (van der Linde et al., 2011). CHD can be classified into broad types: cyanotic heart disease, left-sided obstruction defect and septation defects (Bruneau, 2008). Among CHDs, atrial and ventricular septal defects are the most common subtypes in most districts of the world (van der Linde et al., 2011). A heart development program is precisely and spatially regulated by signal molecules (van den Akker et al., 2012; Kruithof et al., 2012; Schulz and Yutzey, 2004). The regulation of heart development includes a conserved tissue specific transcriptional factor modulation network which is required for morphogenesis of the atrioventricular septum (Olson, 2006; Moorman and Christoffels, 2003). During heart development, any disruption of this programing will cause embryonic lethality or cardiac defect (Bruneau, 2013; Srivastava and Olson, 2000). Nowadays, remarkable progress has made in CHD prevention, diagnosis and cardiac surgery (Killen et al., 2014; Oster et al., 2013; Guleserian, 2011). However, the pathogenic mechanism of CHD has not been well elucidated yet.

Several classic studies have indicated that the etiology of CHD may be attributed to both genetic and environmental factors (Richards and Garg, 2010; Shieh et al., 2012). Some CHDs are characterized with chromosomal deletions or abnormal chromosomal number, and 50% of children who are born with Trisomy 21 (Down syndrome) are affected by cardiac defects (Richards and Garg, 2010). In the past decades, breakthroughs in molecular genetic technology led to better understanding of the key pathways which played important roles during cardiogenesis, such as calcineurin signaling, Jagged-1/Notch signaling, bone morphogenetic protein (BMP) signaling and transforming growth factor (TGF)-β signaling pathways (Schulz and Yutzey, 2004; Krebs et al., 2000; Wang et al., 2011; Anderson and Gibbons, 2007). Moreover, an increasing evidence suggested that a single gene mutation was

Abbreviations: CHD, congenital heart disease; ASD, atrial septal defect; BMP, bone morphogenetic protein; GO, gene ontology; DEGs, differentially expressed genes; FDR, false discovery rate.

* Correspondence to: Z. Hou, Department of Central Laboratory, Yan'an Affiliated Hospital of Kunming Medical University, No. 245 East of Renmin Road, Kunming, Yunnan 650051, People's Republic of China.
** Correspondence to: L. Jiang, Department of Cardiovascular Surgery, Yan'an Affiliated Hospital of Kunming Medical University, No. 245 East of Renmin Road, Kunming, Yunnan 650051, People's Republic of China.
E-mail addresses: hzl579@163.com (Z. Hou), jianglihong_yayy@163.com (L. Jiang).
[1] These authors contributed equally to these work.

http://dx.doi.org/10.1016/j.gene.2015.09.016
0378-1119/

presented in patients with CHD. The transcription factors including GATA4, NKX2-5, dHAND, TFAP2 and TBX5 are crucial genes which are required for early heart development (Wessels and Willems, 2010; Marin-Garcia, 2009). Studies have shown that mutations in these genes give rise to severe cardiac defect such as septal defects (GATA4, NKX2-5, TBX5), conduction defects (NKX2-5), right ventricular hypoplasia (HAND2), patent ductus arteriosus in Char syndrome (TFAP2B) and Holt–Oram syndrome (TBX5) (McCulley and Black, 2012; Fahed et al., 2013). Risk loci for CHD were identified by genome-wide association study (Hu et al., 2013). However, the dysregulation of CHD on transcription level is still unclear. In order to identify the dysregulated genes which may cause cardiac defect, we performed comparative transcriptome analysis between normal and atrial septal defect (ASD) patients. The mRNAs isolated from tissues of normal and ASD were transcribed into cDNAs, and applied to Illumina RNA sequencing (RNA-seq). The results of RNA-seq were analyzed by a series of bioinformatic methods including mapping, gene differential expression analysis, gene ontology (GO) and pathway analysis. Our study provided evidences of understanding pathogenic mechanism of ASD and resources for validation of CHD genomic studies.

2. Materials and methods

The study and protocol were reviewed and approved by Bioethics Committee of Yan'an Affiliated Hospital of Kunming Medical University. The study was carried out in accordance with the approved guidelines. Written informed consents have been given from all patients and their legal guardian. All procedures in this study were performed in compliance with Helsinki Declaration and national laws.

2.1. Tissue sample collection

Atrial septal defect tissue samples were obtained from ASD patients who underwent atrial septal defect repair surgery in Yan'an Affiliated Hospital of Kunming Medical University. The samples from ASD were the rims of the hole in the atrial septum which were excised and cast off before the tissue patch fixed in. Experimental subject enrollment criteria included: age of the infants between 6 and 12 months; ostium secundum type of atrial septal defect diagnosed by echocardiography; excluding ostium primum ASD and infants with other common development defects. Atrial septum tissue samples from normal control were obtained from pregnant women who underwent induced abortion in Yan'an Affiliated Hospital of Kunming Medical University. The gestational ages were between 30 and 33 weeks who underwent voluntary abortion due to some private reasons. Three ASD and three normal control subjects were included in this study. For collection of mouse embryo atrial septum tissue, wide-type C57BL/6 strain mouse embryos were harvested at defined stages from timed mating pairs. By employing a stereo microscope, heart tubes were collected at E8.5, and atrial septum of the embryos were microdissected including E10.5, E12.5, E13.5, E14.5, E15.5, E17.5, E21.5 and adult mice. The E21.5 embryos were newborn mice, and adult mice were 6 months old. Five independent samples were harvested for each mouse embryo development stage, and the tissue samples were frozen in liquid nitrogen for RNA extraction.

2.2. RNA isolation and quality control

To extract RNA, frozen tissues were disrupted using liquid nitrogen with mortar and pestle. The tissue powder was suspended in TRIzol Reagent (Life technologies, USA) according to the protocol from the manufacturer. The concentration of each sample was determined by NanoDrop 2000 (Thermo Scientific, USA). The quality was assessed by the Agilent 2200 (Agilent, USA).

2.3. Whole transcriptome library preparation and deep sequencing

The sequencing RNA library of each sample was prepared using TruSeq RNA Library Preparation Kit v2 following the protocol provided by the manufacturer (Illumina, San Diego, CA). Briefly, polyA containing mRNA was purified from 1 μg of total RNA using oligo-dT attached magnetic beads, and fragmentation and priming were done during the second round purification. The primed mRNA fragments were reverse-transcribed into double stranded cDNA. Then, the 3′ overhangs of the cDNA were converted into blunt ends for A-tailing. By priming with single "A" nucleotide, T-tailed adaptor was ligated to the fragment. The ligated products were purified and selected for enrichment by PCR. The concentration of the enriched library was adjusted for deep sequencing by Illumina Hiseq2000 Sequencer.

2.4. Filtering the raw reads and mapping

All adaptors and ribosome RNA sequences were excluded from the raw reads. For each sample, reads with quality score of ≥Q20 passed the filtering. The reads ≥50 bp which passed filtering were used for mapping. Filtered sequencing reads were aligned against the reference genome using the spliced mapping algorithm in the Tophat package. This algorithm allows separating a read that cannot be mapped as a whole and map different parts of the read separately to the genome, which is useful for mapping to eukaryotic transcriptomes. And the parameters allow two mismatches in a mapping event and multi-hits ≤1 in each read. The result of sequence alignment was saved in BAM format. The reference genome is ensembl Homo_sapiens.

2.5. Gene expression and differential gene analysis

Sequence alignments after Tophat were further processed with the Cufflinks software (version 2.1.1) to assemble transcripts, quantify the expression level and analyze differently expressed genes. The gene expression level was quantified using FPKM value. The equation to quantify the FPKM value is as follows: FPKM (definition): Fragments Per Kilobase of exon model per Million mapped reads. For ASD and normal control samples, the Cuffdiff module of the Cufflinks software was used to analyze differential expressions. To choose differentially expressed genes, the fold change and Fisher-test were used. We selected the differentially expressed genes by the following criteria: p value <0.05, FDR <0.05, and fold change ≥2 (Anders and Huber, 2010).

2.6. Gene ontology (GO) and pathway analysis

GO analysis was applied to analyze the main function of DEGs according to the Gene Ontology which is the key functional classification of NCBI (Anon, 2006; Ashburner et al., 2000). Likewise, pathway analysis was used to find out the significant pathway of DEGs according to KEGG and Biocarta and Reactome (Kanehisa and Goto, 2000; Joshi-Tope et al., 2005). The method of GO and pathway analysis was described previously (Wang et al., 2015). Briefly, GO analysis of DEGs was performed by employing DAVID gene annotation tool (Dennis et al., 2003; Huang da et al., 2009). Statistical analysis of GO terms was done by Fisher's exact test and χ^2test, and the false discovery rate (FDR) was calculated to correct the p value, the smaller the FDR, the small the error in judging the p value (Benjamini, 1995). The significant GO terms were defined as p value <0.05 and FDR <0.05. Still, we turned to the Fisher's exact test and χ^2test to select the significant pathway using DAVID annotation tool, and the threshold of significance was defined by p value and FDR. The significant pathway was identified by p value <0.05 and FDR <0.05.