Gene 574 (2015) 53–60

Contents lists available at ScienceDirect

Gene

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

Research paper

Deep sequencing-generated modules demonstrate coherent expression patterns for various cardiac diseases

Seong-Eui Hong, Kyoung Jin Nho, Hong Ki Song, Do Han Kim *

College of Life Sciences and Systems Biology Research Center, Gwangju Institute of Science and Technology (GIST), 123 Cheomdangwagi-ro, Buk-gu, Gwangju 500-712, Republic of Korea

ARTICLE INFO

Article history:
Received 5 March 2015
Received in revised form 19 June 2015
Accepted 24 July 2015
Available online 29 July 2015

Keywords:
Heart-enriched gene
Gene module
Next-generation sequencing
Cardiac function

ABSTRACT

As sequencing technology rapidly develops, gene annotations have also become increasingly sophisticated with incorporation of information regarding the temporal–spatial context of alternative splicing patterns, developmental stages, and tissue specificity. The present study aimed to identify the heart-enriched genes based on next-generation sequencing data and to investigate the gene modules demonstrating coherent expression patterns for various cardiac disease-related perturbations. Seven gene modules, including 382 heart-enriched genes, were identified. At least two modules containing differentially expressed genes were experimentally confirmed to be highly sensitive to various cardiac diseases. Transcription factors regulating the gene modules were then analyzed based on knowledgebase information; the expression of eight transcription factors changed significantly during pressure-overload cardiac hypertrophy, suggesting possible regulation of the modules by the identified transcription factors. Collectively, our results contribute to the classification of heart-enriched genes and their modules and would aid in identification of the transcription factors involved in cardiac pathogenesis in the future.

1. Introduction

The rapid development of sequencing technologies has led to the accumulation of large amounts of high-throughput transcriptomic data. In recent decades, the microarray method has made a great contribution to the analysis of transcriptomes of various diseases. However, current state-of-the-art sequencing techniques, known as next-generation sequencing (NGS), have brought about a more comprehensive understanding of transcriptomes, enabling the simultaneous analysis of both quantitative (e.g., gene and non-coding RNA) and qualitative signatures (e.g., alternative splicing) (Werner, 2010; Derrien et al., 2012; Song et al., 2012), while facilitating the identification of novel gene structures (e.g., novel transcripts and exons) by *de novo* analysis (Chen et al., 2013; Lee et al., 2011). With an increasing understanding of transcriptomes, it is possible to assess the more sophisticated gene annotations in terms of the temporal–spatial regulation of genes with respect to disease status, developmental stages, and tissue specificity (Song et al., 2012; DeLaughter et al., 2013; Wang et al., 2012; Brawand et al., 2011).

Gene annotation according to the tissue specificity of genes has been an important challenge in understanding the function of the organs and in the design of tissue-specific targeted drugs. As a result, many previous studies have focused on the analysis of tissue-specific transcriptomes (Sun et al., 2013; Cotney et al., 2012; Song et al., 2013) and on the construction of databases containing information related to tissue specificities (Hong et al., 2006; Barshir et al., 2013; Xiao et al., 2010). In the context of cardiac cells, several research groups have focused their work on identifying heart-specific genes and their modules based on the transcriptomes discovered using microarray and expressed sequence tag (EST) techniques (Johnson et al., 2005; Qian et al., 2005). We have also investigated heart-specific genes based on the EST data deposited in the UniGene database (Hong et al., 2006). The present study employs NGS data for the identification of heart-enriched genes.

In this study, 382 heart-enriched genes and seven modules that demonstrate the coherent expression patterns for various disease perturbations in the heart were identified. Given the modules, the significantly enriched transcription factors in the modules were investigated, and the differential expression of transcription factors and the genes involved in the pathogenesis of cardiac hypertrophy were experimentally confirmed.

2. Material and methods

2.1. NGS data analysis

For the identification of tissue-enriched genes, GSE30352 was downloaded (Brawand et al., 2011). For mapping onto mouse genome (UCSC mm9 version), Tophat was applied to the fastq files for six different tissues (Kim et al., 2013) and Samtools further converted into sam files (Li et al., 2009). For the sam files, HTSeq was employed to calculate the number of reads for the genes annotated by RefSeq (Pruitt et al., 2014) and then, DESeq was finally used to analyze the differentially expressed genes in one tissue over other tissues ($q < 0.05$, $|\log_2(\text{fold change})| \geq 1$) (Anders et al., 2013). The genes showing the abundant expression in heart than other tissues were collected as heart-enriched genes. For the expression profiles for the heart-enriched genes, the multiple NGS data (i.e., E-MTAB-727 Song et al., 2012, GSE29446 Lee et al., 2011, GSE30943 Sakabe et al., 2012, GSE31839, GSE34274 Delgado-

http://dx.doi.org/10.1016/j.gene.2015.07.080
0378-1119/

Olguin et al., 2012, GSE35350 Mitchell et al., 2006, GSE37076 Solovei et al., 2013, GSE39911 Wang et al., 2012 and GSE48043 Flynn et al., 2013) were analyzed using the same pipeline, "Tophat–Samtools–HTSeq–DESeq".

2.2. Construction of gene modules

Based on the expression profiles from the multiple NGS data, the heart-enriched genes were clustered using CLICK using the default parameter (Sharan et al., 2003).

2.3. Pathway analysis

Pathways regulated by the genes were analyzed using Ingenuity Pathway Analysis (IPA) (Ingenuity Pathway Analysis, nd). To assess the functions, "Diseases and Functions" of IPA was analyzed for the genes in each module. The significance of $p < 0.05$ was adopted.

2.4. Parametric analysis of gene set enrichment (PAGE) analysis

The representative expression levels were measured by modified PAGE algorithm as previously reported (Kim and Volsky, 2005). The Z-score was expressed as follows:

$$Z = \frac{\left(\overline{FC}^{module} - \overline{FC}^{overall}\right) \times \sqrt{n}}{\delta^{overall}}$$

where $\overline{FC}^{module}$ and $\overline{FC}^{overall}$ are the averages of the log-scaled fold changes of the genes in a module and overall, respectively, n is the number of genes in a module, and $\delta^{overall}$ is the standard deviation of the overall fold changes.

2.5. Ethics statement

All animal experiments were approved by the Gwangju Institute of Science and Technology Animal Care and Use Committee (2014–55).

2.6. Transverse aortic constriction operation

8 week old male (C57BL/6J) mice (body weight 28–33 g) purchased from Samtako Korea were used in all studies. Pressure-overload cardiac hypertrophy was induced by transverse aortic constriction (TAC) operation under anesthesia with intraperitoneal injection of avertin, 2-2-2 tribromoethanol (Sigma, St. Louis, MO) dissolved in tert-amyl alcohol (Sigma, St. Louis, MO). The procedure of operation was followed as previously described (Song et al., 2012). As a control group, sham operation (same procedure except for tying) was done. 1 week after operation, mice were euthanized by cervical dislocation, and hearts were removed. The visual parts of aorta were removed from the heart as far as possible. The trimmed hearts were stored in deep freezer at −80 °C before protein and RNA extraction.

2.7. RNA isolation and qRT-PCR

Mouse hearts were snap frozen in liquid nitrogen, stores at −80 °C, and homogenized in liquid nitrogen using a mortar and pestle. Approximately 450–700 mg of grinded whole mouse heart was used for extraction of total RNA with 1 ml Trizol Reagent® (Invitrogen, Carlsbad, CA) following the manufacturer's instructions. First-strand cDNA was synthesized from 2 μg of total RNA with Random hexamer using Omniscript® reverse transcription (Qiagen, Valencia, CA) according to the manufacturer's instruction. Briefly, qRT-PCR assays were performed using TOPreal™ qPCR premix (Enzynomics, Korea) under the following two-step conditions: denaturation at 95 °C for 15 s followed by annealing and extension at 60 °C for 40 s, for a total of 40 cycles. The 18S transcript was used as an endogenous reference to assess the relative level of mRNA transcript.

2.8. Visualization

The gene modules and expression profiles across the multiple perturbations were illustrated using Expander (Ulitsky et al., 2010).

3. Results

3.1. Identification of heart-enriched genes

In order to identify the genes highly enriched in cardiac cells, we analyzed the existing high-throughput data (GSE30352 Brawand et al., 2011) examining tissue distribution. The transcriptomic data included six different mouse tissues such as brain, cerebellum, heart, kidney, liver, and testis. Each tissue-enriched gene was identified using the pipeline "Tophat–Samtools–HTSeq–DESeq" and the RefSeq database was used for gene annotations. Among a total of 26,674 genes tested, the genes showing significantly higher expression levels in a particular tissue (q-value < 0.05, $|\log_2$fold change$| \geq 1$) in comparison with other tissues were classified as tissue-enriched.

As a result, 382 heart-enriched genes and approximately 2400 other tissue-enriched genes were identified (Table S1). Among the 382 heart-enriched genes, many genes involved in metabolic pathways, such as oxidative phosphorylation in mitochondria, were found to be significantly enriched ($q < 0.01$). This was particularly distinguishable from other functions enriched in other tissue-enriched genes (Fig. 1). The enrichment of genes from the mitochondria in the heart-enriched genes might be a result of the energy demands of the beating heart (Stanley and Hoppel, 2000; Murray et al., 2007). In addition, 79 of these genes are known to be associated with cardiovascular diseases ($q = 8.91 \times 10^{-10}$) and 83 genes are involved in cardiovascular system development and function ($q = 7.88 \times 10^{-9}$) based on IPA (Table S2).

3.2. Construction of heart-enriched gene modules

We then investigated the gene modules demonstrating coherent expression patterns across multiple transcriptomes for the heart-enriched genes. NGS data on the 16 disease-related perturbations in the heart or muscle from 9 NGS datasets were analyzed with the same pipeline as used for the identification of heart-enriched genes (Tophat–Samtools–HTSeq–DESeq). We constructed the expression patterns of 382 heart-enriched genes for the perturbations. The expression profiles were further clustered using the CLICK algorithm (Sharan and Shamir, 2000). As a result, seven gene modules demonstrating homogeneous expressions were identified (Fig. 2); detailed information on the modules is provided in Table S3.

In order to understand the characteristics of each module, we assessed the enriched functions for the modules using IPA. We found that 29 functions were significantly enriched in at least one module (Fisher's exact test, $q < 0.001$). The representative functions are shown in Fig. 3A. Modules 1, 2 and 5 were closely related to the various functions noted in the case of the heart. The genes in module 5 were most likely also involved in various cardiac functions. Of the genes in module 5, 36 genes are known to be related to numerous cardiac diseases based on the knowledgebase information from IPA (Table 1). In particular, cardiomyopathy was closely associated with seven of the heart-enriched genes (*Cav1*, *Grk5*, *Lama4*, *Lims1*, *Mybp2*, *Sgcd*, and *Vcl*) in module 5 ($p = 2.19 \times 10^{-6}$). The genes in module 5 are also known to be linked to cell morphology ($p = 1.54 \times 10^{-10}$), particularly, cell spreading. A total of 10 heart-enriched genes included in this module (*Flnc*, *Hspg2*, *Itga5*, *Lama2*, *Parvb*, *Pecam1*, *Sorbs1*, *Tgfb1l1*, *Vcl*, and *Vim*) are known to be involved in cell spreading.