Gene 552 (2014) 75–80

Contents lists available at ScienceDirect

Gene

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

Methods paper

Digital gene expression analysis of the pathogenesis and therapeutic mechanisms of ligustrazine and puerarin in rat atherosclerosis

Rao Fu [a,1], Yuting Zhang [a,1], Yiran Guo [a,*], Yanling Zhang [b], Ying Xu [a], Fang Chen [a,*]

[a] College of Life Sciences, Sichuan University, National and Local Joint Engineering Laboratory for Energy Plant Bio-oil Production and Application, Key Laboratory of Bio-resource and Eco-environment, Ministry of Education, Chengdu, PR China
[b] Beijing University of Chinese Medicine, Beijing, PR China

ARTICLE INFO

Article history:
Received 2 July 2014
Received in revised form 18 August 2014
Accepted 8 September 2014
Available online 9 September 2014

Keywords:
Digital gene expression
Pathogenesis
Ligustrazine
Puerarin
Atherosclerosis

ABSTRACT

Atherosclerosis (AS) is the leading cause of death in modern societies. Active substance from Traditional Chinese Medicine has been used for the treatment of AS, such as ligustrazine and puerarin. However, the pathogenesis of AS and the curative mechanisms of ligustrazine and puerarin stay unclear. In this work, we attempted to figure out these questions using a rat AS model and digital gene expression (DGE) system. Our results showed that DGE sequencing outcomes were high quality and reproductively. Differentially expressed genes were obtained from different comparisons. The Gene Ontology (GO) analysis revealed that mainly enriched GO terms due to the drug treatment were the same as those obtained from the control group vs. the AS model group. Pathway analysis indicated that metabolic pathways, oxidative phosphorylation, and PPAR single pathways were enriched in all comparisons. Our work provided a comprehensive basis for a better understanding of the pathogenesis of AS and the curative mechanisms of ligustrazine and puerarin.

1. Introduction

The rhizome of *Ligusticum chuanxiong* Hort. has long been used for treating blood stasis in Traditional Chinese Medicine (TCM). Ligustrazine was identified as an active ingredient (Sutter and Wang, 1993), which has been reported to be effective for the treatment of a variety of vascular diseases in clinical. Puerarin is the most abundant isoflavone-C-glucoside compound isolated from *Pueraria lobata*, and has been widely used for the treatment of cardiovascular, neurological and hyperglycemic disorders (Shi et al., 2002). It has been founded that both of ligustrazine and puerarin would be benefit to cardiovascular diseases, including atherosclerosis (AS) (Yan et al., 2006; Jiang et al., 2011). AS is the leading cause of death with a universal incidence in modern societies, which is not only a disease in its own right, but also a process that is the primary contributor to the pathogenesis of myocardial and cerebral infarction, gangrene, and loss of function in the extremities (Wu et al., 2009). Although a growing body of evidences indicated that AS is a chronic inflammatory disease (Zapolska-Downar and Zapolski-Downar, 2002; Lusis, 2000), and hypercholesterolemia is an important risk factor for the development and progression of AS disease (Subramanian et al., 2003). But, the pathogenesis of AS is quite complicated so far, even though a lot of work has been done to investigate the exact pathogenesis of AS, unfortunately, which stays unclear.

Recently, high-thought sequencing technology has been adopted for transcriptome analysis, and has dramatically improved the efficiency and speed of gene discovery (Ansorge, 2009). Digital gene expression (DGE) developed by Illumina (formerly Solexa sequencing) allows millions of short RNAs and differently expressed genes (DEGs) to be identified in a sample without prior annotations (Bennett et al., 2005). DGE has been widely used for the investigation of DEGs and exhibited high accuracy, stability and repeatability that has been proved by RT-qPCR in many reports (Xiao et al., 2010; Tang et al., 2011).

In the present study, we firstly used DGE technology to clarify the pathogenesis of AS in a rat model and the mechanisms of ligustrazine and puerarin in the treatment of AS. The method of administration of vitamin D_3 and cholesterol was chosen to establish the animal model. The transcriptome profiles of rat thoracic aortas were investigated. DEGs were founded and analyzed comprehensively. Our results may serve as a basis for future research in pathology of AS and molecular mechanisms of related promoting blood circulation and removing blood stasis drugs in the treatment of AS.

Abbreviations: AS, atherosclerosis; DGE, digital gene expression; GO, Gene Ontology; TCM, Traditional Chinese Medicine; DEGs, differently expressed genes; PEPCK, phosphoenolpyruvate carboxykinase; ROS, reactive oxygen species; TPM, number of transcripts per million clean tags; FDR, false discovery rate.

* Corresponding authors at: College of Life Science, Sichuan University, No. 24 South Section 1, Yihuan Road, Chengdu, PR China.
E-mail addresses: ran@scu.edu.cn (Y. Guo), fangchen1112@gmail.com (F. Chen).

[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.gene.2014.09.015
0378-1119/

2. Results

2.1. Analysis of DGE libraries

Ten DGE libraries have been constructed and sequenced, including: control (A1 and A2), AS model (C1 and C2), ligustrazine treatment (E1, E2 and E3) and puerarin treatment (F1, F2 and F3). The major characteristics of each library were summarized in Table 1. We got 3.5–3.7 million raw tags, and after being filtered, approximately 3.3–3.5 million clean tags corresponding to 0.1–0.12 million distinct tag numbers were obtained. The distribution of total and distinct clean tags over the different tag abundance categories showed highly similar patterns for all DGE libraries (Figs. 1 and S1). More than 70% of the total clean tags had a copy number higher than 100, whereas less than 4.3% of distinct tags had a copy number higher than 100. Tags with a copy number between 2 and 5 had a large scale distribution of distinct tags. These results demonstrated that the majority of mRNA were expressed at a low level, only a small proportion of mRNA made a difference. To further investigate the reproducibility of DGE library sequencing; we performed intra-group correlation analyses, including A1 vs. A2, C1 vs. C2, E1 vs. E2, E1 vs. E3, E2 vs. E3, F1 vs. F2, F1 vs. F3 and F2 vs. F3. The Pearson correlations were 0.933, 0.984, 0.986, 0.962, 0.959, 0.948, 0.929 and 0.943, respectively. These results suggested that sequencing results have high reproducibility. Sequencing data saturation refers to the status that no more new unique tags can be detected with the increases of the number of total tags. Our results showed that the number of detected genes was almost saturated when the total tag number reached 2 million or higher (Fig. S2). Above all, ten DGE libraries of our research were reliable and can be employed for further analysis.

2.2. Analysis of tag mapping

We mapped tag sequences of the ten DGE libraries to the reference database consisted of *Rattus norvegicus* genome and genes. Approximately 82.89%–85.39% of the clean tags corresponding to 53.84%–60.60% of distinct tags were mapped to a gene in the reference database. 50.23%–56.43% of the distinct clean tags were mapped unambiguously to the UniGene database, 16.74%–19.96% of the distinct clean tags could be mapped to genome, while 8.72%–11.12% of the clean tags were unknown tags (Table 1). The number of unambiguous tags for each gene was normalized to TPM to accurately assess the gene expression level. The averaged TPM of each gene in a group was calculated for DEGs, and GO and pathway enrichment analysis.

2.3. Identification of DEGs

To identify genes showing significant changes between groups, we analyzed the DEGs between each two groups (Fig. 2). For the control vs. the AS model (Table S1), the Pearson correlation coefficient of the two libraries was 0.84, suggesting that the AS induced the changes of gene expression profile of rat thoracic aorta which should be related to the pathogenesis of AS. DEGs generated by AS rat treatment with ligustrazine and puerarin vs. AS model were listed at Tables S2 and S3, respectively. Treatment of ligustrazine showed fewer DEGs than puerarin which may be related to its more concentrated targets or the dose we used.

2.4. GO enrichment analysis

To understand the roles of these DEGs between different groups, all the DEGs were mapped to terms in GO database and compared with the whole genome background. DEGs due to the occurrence of AS can be categorized into a total of 130 GO terms (cellular component: 23 terms; molecular function: 26 terms; biological process: 81 terms) (Table S4). In each of the three categories of the GO classification, “cytoplasm”, “cytoplasmic part”, “catalytic activity”, “binding”, “metabolic process” and “cellular metabolic process” terms are dominant. Interestingly, it is already known that metabolic syndrome is the main risk factor of AS. After the treatment of ligustrazine, as compared with the AS model group, the DEGs could be enriched to 20 GO terms (cellular component: 4 terms; molecular function: 1 term; biological process: 16 terms) (Table S5). “Cytoplasm”, “cytoplasmic part”, “binding” and “response to stimulus” terms were dominant in each of the three categories. Meanwhile, the genes changed after the administration of puerarin could be categorized into a total of 119 GO terms (cellular component: 27 terms; molecular function: 20 terms; biological process: 72 terms) (Table S6). “Membrane-bounded organelle”, “cytoplasm”, “binding”, “oxidoreductase activity”, “metabolic process”, “cellular metabolic process” and “response to stimulus” were dominant.

Table 1
Categorization and abundance of tags. Clean tags are the tags after filtering dirty tags from raw tags. Distinct tags are different kinds of tags. Unambiguous tags are the remainder clean tags after removing tags mapped to reference sequences from multiple genes.

Summary		A1	A2	C1	C2	E1	E2	E3	F1	F2	F3
Raw data	Total	3,648,250	3,559,605	3,456,242	3,740,361	3,642,081	3,642,004	3,748,108	3,462,571	3,688,930	3,554,795
	Distinct tag	332,869	306,767	323,478	314,739	298,124	301,603	356,338	276,808	275,947	297,681
Clean tag	Total number	3,435,209	3,358,392	3,244,518	3,532,269	3,445,723	3,445,706	3,511,398	3,286,259	3,510,337	3,360,185
	Distinct tag number	120,600	106,414	112,597	107,785	102,785	106,281	120,509	101,210	98,216	103,873
All tag mapping to gene	Total number	2,872,076	2,823,322	2,689,252	2,969,437	2,929,197	2,919,235	2,946,126	2,786,388	2,983,044	2,869,340
	Total % of clean tag	83.61%	84.07%	82.89%	84.07%	85.01%	84.72%	83.90%	84.79%	84.98%	85.39%
	Distinct tag number	66,368	59,905	60,621	58,529	59,776	60,442	66,524	60,266	59,522	62,748
	Distinct tag % of clean tag	55.03%	56.29%	53.84%	54.30%	58.16%	56.87%	55.20%	59.55%	60.60%	60.41%
Unambiguous tag mapping to gene	Total number	2,611,595	2,587,841	2,504,087	2,759,562	2,704,590	2,695,714	2,704,048	2,542,554	2,714,987	2,604,794
	Total % of clean tag	76.02%	77.06%	77.18%	78.12%	78.49%	78.23%	77.01%	77.37%	77.34%	77.52%
	Distinct tag number	62,006	55,832	56,563	54,372	55,631	56,223	62,171	56,401	55,427	58,553
	Distinct tag % of clean tag	51.41%	52.47%	50.23%	50.44%	54.12%	52.90%	51.59%	55.73%	56.43%	56.37%
All tag-mapped genes	Number	24,033	22,667	23,736	23,238	22,682	23,184	24,012	22,746	22,236	23,035
	% of ref genes	35.88%	33.84%	35.44%	34.70%	33.87%	34.62%	35.85%	33.96%	33.20%	34.39%
Unambiguous tag-mapped genes	Number	22,235	20,911	21,966	21,559	20,941	21,419	22,196	20,997	20,504	21,221
	% of ref genes	33.20%	31.22%	32.80%	32.19%	31.27%	31.98%	33.14%	31.35%	30.61%	31.68%
Mapping to genome	Total number	234,624	223,580	194,505	192,906	192,658	196,124	219,386	208,277	216,489	197,689
	Total % of clean tag	6.83%	6.66%	5.99%	5.46%	5.59%	5.69%	6.25%	6.34%	6.17%	5.88%
	Distinct tag number	23,856	19,977	22,471	21,023	18,262	19,700	23,188	17,852	16,437	17,622
	Distinct tag % of clean tag	19.78%	18.77%	19.96%	19.50%	17.77%	18.54%	19.24%	17.64%	16.74%	16.96%
Unknown tag	Total number	328,509	311,490	360,761	369,926	323,868	330,347	345,886	291,594	310,804	293,156
	Total % of clean tag	9.56%	9.27%	11.12%	10.47%	9.40%	9.59%	9.85%	8.87%	8.85%	8.72%
	Distinct tag number	30,376	26,532	29,505	28,233	24,747	26,139	30,797	23,092	22,257	23,503
	Distinct tag % of clean tag	25.19%	24.93%	26.20%	26.19%	24.08%	24.59%	25.56%	22.82%	22.66%	22.63%