Dear Author,

Here are the final proofs of your article. Please check the proofs carefully.

All communications with regard to the proof should be sent to bmcproductionteam2@spi-global.com.

Please note that at this stage you should only be checking for errors introduced during the production process. Please pay particular attention to the following when checking the proof:

- Author names. Check that each author name is spelled correctly, and that names appear in the correct order of first name followed by family name. This will ensure that the names will be indexed correctly (for example if the author’s name is ‘Jane Patel’, she will be cited as ‘Patel, J.’).
- Affiliations. Check that all authors are cited with the correct affiliations, that the author who will receive correspondence has been identified with an asterisk (*), and that all equal contributors have been identified with a dagger sign (†).
- Ensure that the main text is complete.
- Check that figures, tables and their legends are included and in the correct order.
- Look to see that queries that were raised during copy-editing or typesetting have been resolved.
- Confirm that all web links are correct and working.
- Ensure that special characters and equations are displaying correctly.
- Check that additional or supplementary files can be opened and are correct.

Changes in scientific content cannot be made at this stage unless the request has already been approved. This includes changes to title or authorship, new results, or corrected values.

How to return your corrections

Returning your corrections via online submission:
- Please provide details of your corrections in the online correction form. Always indicate the line number to which the correction refers.

Returning your corrections via email:
- Annotate the proof PDF with your corrections.
- Send it as an email attachment to: bmcproductionteam2@spi-global.com.
- Remember to include the journal title, manuscript number, and your name when sending your response via email.

Note: in order to ensure timely publication, if we do not hear from you within 48 hours we may take the decision to sign-off the article on your behalf and proceed to publication.

After you have submitted your corrections, you will receive email notification from our production team that your article has been published in the final version. All changes at this stage are final. We will not be able to make any further changes after publication.

Kind regards,

BioMed Central Production Team 2
Identification of transcription factors and single nucleotide polymorphisms of \textit{Lrh1} and its homologous genes in \textit{Lrh1}-knockout pancreas of mice

Maochun Tang$^{1\dagger}$, Li Cheng$^{2\dagger}$, Rongrong Jia$^1$, Lei Qiu$^1$, Hua Liu$^1$, Shu Zhou$^1$, Xiuying Ma$^1$, Guoyong Hu$^2$, Xingpeng Wang$^2*$ and Yan Zhao$^1*$

\textbf{Abstract}

\textbf{Background:} To identify transcription factors (TFs) and single nucleotide polymorphisms (SNPs) of \textit{Lrh1} (also named \textit{Nr5a2}) and its homologous genes in \textit{Lrh1}-knockout pancreas of mice.

\textbf{Methods:} The RNA-Seq data GSE34030 were downloaded from Gene Expression Omnibus (GEO) database, including 2 \textit{Lrh1} pancreas knockout samples and 2 wild type samples. All reads were processed through TopHat and Cufflinks package to calculate gene-expression level. Then, the differentially expressed genes (DEGs) were identified via non-parametric algorithm (NOISeq) methods in R package, of which the homology genes of \textit{Lrh1} were identified via BLASTN analysis. Furthermore, the TFs of \textit{Lrh1} and its homologous genes were selected based on TRANSFAC database. Additionally, the SNPs were analyzed via SAM tool to record the locations of mutant sites.

\textbf{Results:} Total 15683 DEGs were identified, of which 23 was \textit{Lrh1} homology genes (3 up-regulated and 20 down-regulated). Fetoprotein TF (FTF) was the only TF of \textit{Lrh1} identified and the promoter-binding factor of FTF was \textit{CYP7A}. The SNP annotations of \textit{Lrh1} homologous genes showed that 92% of the mutation sites were occurred in intron and upstream. Three SNPs of \textit{Lrh1} were located in intron, while 1819 SNPs of \textit{Phkb} were located in intron and 1343 SNPs were located in the upstream region.

\textbf{Conclusion:} FTF combined with CYP7A might play an important role in \textit{Lrh1} regulated pancreas-specific transcriptional network. Furthermore, the SNPs analysis of \textit{Lrh1} and its homology genes provided the candidate mutant sites that might affect the \textit{Lrh1}-related production and secretion of pancreatic fluid.

\textbf{Keyword:} \textit{Lrh1}-knockout pancreas, RNA-Seq, \textit{Lrh1} homologous gene, Transcription factor, Single nucleotide polymorphisms

\textbf{Background}

The pancreas is an endocrine gland, producing insulin, glucagon, somatostatin, and pancreatic polypeptide, and also an exocrine gland, accounting for more than 98% of pancreatic gland and secreting pancreatic juice containing digestive enzymes [1]. These digestive enzymes help to further break down the carbohydrates, proteins and lipids in the chime and thus support the absorption and digestion of nutrition in small intestine [2]. In the past decades, many research have focused on target genes and transcription factors (TFs) involved in the exocrine pancreas-specific transcriptional networks which are required for the production and secretion of pancreatic fluid that helps out the digestive system. Currently, many exocrine pancreas-specific genes and transcription factors have been identified, which may promote the understanding of the effect of exocrine pancreas on digestive system.
Liver receptor homolog-1 (Lrh1; also called Nr5A2) is a nuclear receptor of ligand-activated transcription factors in liver by binding as a monomer to DNA sequence elements with the consensus sequence 5'-Py-CAAGGPyCPu-3' [3]. It has been suggested that Lrh1 is progressively expressed in both the endocrine and exocrine pancreas [4]. Baquié M et al. [5] have found that Lrh1 is expressed in human islets and protects β-cells against stress-induced apoptosis that may be mediated via the increased glucocorticoid production that blunts the pro-inflammatory response of islets. Meanwhile, Fayard E et al. [6] have demonstrated that both Lrh1 and CEL (encoding carboxyl ester lipase) are co-expressed and confined to the exocrine pancreas. The identification of CEL as an Lrh1-target gene indicates that Lrh1 plays an important role in enterohepatic cholesterol homeostasis associated with the absorption of cholesterol esters and the assembly of lipoproteins by the intestine [7]. Besides, Lrh1 is a downstream target in the PDX-1 (lead to pancreas agenesis) regulatory cascade that is activated only during early stages of pancreas development and that governs pancreatic development, differentiation and function [8]. Recently, the rapid advent of next-generation sequencing has made this technology broadly available for researchers in various molecular and cellular biological fields. Holmstrom SR et al. [9] have determined the cistrome and transcriptome for the nuclear receptor LRH-1 in exocrine pancreas and revealed that Lrh1 directly induces expression of genes encoding digestive enzymes and secretory and mitochondrial proteins based on Chromatin immunoprecipitation (ChIP)-seq and RNA-seq analyses. Besides, Lrh1 cooperates with the pancreas transcription factor 1-L complex (PTF1-L) in regulation of exocrine pancreas-specific gene expression. However, many potential target genes and TFs of Lrh1 based on RNA-seq analysis have not been revealed.

In the present study, we downloaded the raw RNA-seq data of Holmstrom SR et al. deposited in The National Center for Biotechnology Information (NCBI) database, which were analyzed using multiple bioinformatics tools in the purpose of finding specific TFs of Lrh1 and its homologous genes. Additionally, we also annotated the SNPs of Lrh1 and its homologous genes to predict their mutant sites. Our study might improve the understanding of the regulation network of Lrh1-related production and secretion of pancreatic fluid.

Methods
RNA-seq data acquisition
The RNA-seq data was downloaded from NCBI (http://www.ncbi.nlm.nih.gov/) Gene Expression Omnibus (GEO) database (GEO accession: GSE34030 [9]), including 2 Lrh1 pancreas knockout samples and 2 wild type samples. RNA preparations were subjected to the Illumina RNA-seq protocol and the platform was GPL9185.

Data pre-processing, gene expression and homology gene of Lrh1
The raw data were downloaded from SRA (Sequence Read Archive) of NCBI and then converted to fastq reads using fastq-dump program of NCBI SRA Toolkit (–q 64) (http://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=std). Then, these reads were processed through TopHat [10] and Cufflinks [11] package to calculate gene-expression level. All parameters were set up according to the default settings of TopHat and Cufflinks. The DEGs were identified via non-parametric algorithm (NOISeq) methods in R package [12]. The thresholds value was False Discovery Rate (FDR) < 0.001. BLASTN analysis of the selected DEGs was used to identify the homology genes of Lrh1. Homology genes here refer to the paralogous genes which share a high degree of sequence similarity (maximum expectation value was set to e−5) with Lrh1 in mice.

Function annotation of Lrh1 homologous genes
For functional analysis of Lrh1 homologous genes, DAVID (Database for Annotation, Visualization and Integrated Discovery) [15] was performed for Gene Ontology (GO) [16] function and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis.

Transcription factor (TF) of Lrh1 homologous genes
Combined with TRANSFAC database [17], the TFs regulated the transcription of Lrh1 and its homologous genes were identified. Then, the promoter-binding factors regulated via the selected TFs were analyzed based on the website (http://www.nursa.org/molecule.cfm?molType= receptor&molId=5A2).

Screening of SNPs
The fastq reads were mapped to marker sequences using bowtie [18]. And the aligned reads were called using the SAM tool [19]. In order to minimize the risk of false-positive SNP Callings, the threshold value was that ID was “*” with quality > 50, or ID was not “*” with quality > 20. These SNPs were annotated via SnpEff [20] to categorize the effects of variants in genome sequences. The identified SNPs were searched in the dbSNP database to identify diseased SNPs or de novo discovered SNPs.

Results
Identification and homology analysis of differentially expressed genes
After data processing, at FDR < 0.001, a total of 15683 DEGs were identified, including 10994 up-regulated and 4698 down-regulated genes. BLASTN analysis of DEGs showed 23 Lrh1 homology genes. Among them, 3 were up-regulated and 20 were down-regulated (Table 1).
Function and pathway annotation of Lrh1 homologous genes

To determine the function of Lrh1 homologous genes in pancreas, GO enrichment analysis and KEGG pathway enrichment analysis were used to analyze the up- and down-regulated Lrh1 homologous genes. For function and pathway annotation, DEGs were enriched into hexose metabolic process and monosaccharide metabolic process, which were involved into glycometabolism (Figure 1). Meanwhile, KEGG pathway enrichment analysis identified insulin signaling pathway, indicating that the disorders of glycometabolism might be resulted from insulin resistance and/or insulin secretion (Figure 2). PHKB, an Lrh1 homologous gene, participated in GO terms (hexose metabolic process and monosaccharide metabolic process) and KEGG pathway (insulin signaling pathway), was identified.

Potential TFs of Lrh1 homologous genes

Fetoprotein transcription factor (FTF) (ID: T04754) of Lrh1 was the only TF identified based on TRANSFAC database. Meanwhile, the promoter-binding factor of Lrh1 was CYP7A (Cholesterol 7α-hydroxylase).

SNPs of Lrh1 homologous genes

The annotation of SNPs of Lrh1 homologous genes showed that the majority of SNPs were located in intron and upstream, accounting for nearly 92% of all SNPs (Tables 2 and 3). Three SNPs of Lrh1 were distributed in intron. Meanwhile, total 1819 SNPs of Phkb were located in the intron and 1343 SNPs were located in the upstream region of Phkb.

Discussion

In the present study, combined with RNA-seq data of Lrh1-knockout pancreas samples, FTF was the only TF of Lrh1 identified based on TRANSFAC database and may regulate cholesterol catabolism into bile acids by activation of the promoter-binding factor CYP7A. Many literatures have elucidated the function of Lrh1/Nr5a2/FTF/CYP7A via experimental studies [21-25]. FTF is highly expressed in the liver and intestine and is implicated in the regulation of cholesterol, bile acid and steroid hormone homeostasis [26]. Nearly 50% of the body cholesterol is catabolized to bile acids via bile acid biosynthetic pathway, of which cholic acid (hydroxylated at position 12) and chenodeoxycholic acid are the major primary bile acids and play an important role in various physiological processes.

Figure 1 GO functional annotation of Lrh1 and its homology genes. GO terms included Molecular Function (MF) GO-terms (A), Biological Process (BP) GO-terms (B) and Cellular Component (CC) GO-terms (C).
role in cholesterol homeostasis [19]. Chenodeoxycholic acid can repress FTF expression and is a more potent suppressor of HMG-CoA reductase and cholesterol 7α-hydroxylase/CYP7A1 (7α-hydroxylase) than cholic acid [27]. It has been proposed that Lrh1, also known as CYP7A promoter-binding factor, LRH1, or FTF, is required for the transcription of the 7α-hydroxylase gene [19,28]. The small heterodimer partner 1 (SHP) of the nuclear bile acid receptor, FXR (farnesoid X receptor) can dimerize with FTF and diminish its activity on the 7α-hydroxylase promoter [29]. Although Lrh1 has been demonstrated the function in feedback regulation of CYP7A1 expression as part of the FXR-SHP-LRH-1 cascade, in which bile acids can inhibit their own synthesis, the mechanisms have not been well understood. Out C et al. [25] have suggested that CYP7A1 expression is increased rather than decreased under chow-fed conditions in Lrh1-knockdown mice that is coincided with a significant reduction in expression of intestinal Fgf15, a suppressor of CYP7A1. Besides, Noshiro M et al. [30] have suggested that the circadian rhythm of CYP7A is regulated by multiple transcription factors, including DBP, REV-ERβ/δ, LXRa, HNF4α, PPARα, Hepatocyte nuclear factor 4α (HNF4α) and FTF are two major TFs driving CYP7A1 promoter activity in lipid homeostasis. Bochkis IM et al. [31] have shown that prospero-related homeobox (Prox1) directly interacts with both HNF4α and FTF and potently co-represses CYP7A1 transcription.

In the present study, we annotated the SNPs of Lrh1 and its homologous genes, showing that the majority was located in intron and upstream. Quiles Romagosa MÁ [32] has reported that a functional SNP located in Lrh1 promoter is related to Body Mass Index (BMI) and these SNPs might play important roles in the obese phenotype. However, previous researches mostly focused on SNPs associated with pancreatic cancer cell growth.

**Table 2 The number of different type of SNPs of Lrh1 and its homology genes**

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downstream</td>
<td>581</td>
<td>0.052234</td>
</tr>
<tr>
<td>Intron</td>
<td>8251</td>
<td>0.741796</td>
</tr>
<tr>
<td>Non-synonymous-coding</td>
<td>122</td>
<td>0.010968</td>
</tr>
<tr>
<td>Splice-site-donor</td>
<td>2</td>
<td>0.00018</td>
</tr>
<tr>
<td>Start-gained</td>
<td>4</td>
<td>0.00036</td>
</tr>
<tr>
<td>Stop-gained</td>
<td>5</td>
<td>0.00045</td>
</tr>
<tr>
<td>Synonymous-coding</td>
<td>97</td>
<td>0.008721</td>
</tr>
<tr>
<td>Upstream</td>
<td>1975</td>
<td>0.17756</td>
</tr>
<tr>
<td>UTR-3-prime</td>
<td>79</td>
<td>0.007102</td>
</tr>
<tr>
<td>UTR-S-prime</td>
<td>7</td>
<td>0.000629</td>
</tr>
</tbody>
</table>

**Table 3 The annotation of SNPs of Lrh1 and its homology genes**

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Chromo</th>
<th>Number</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zfc3h1</td>
<td>10</td>
<td>4</td>
<td>115385510</td>
</tr>
<tr>
<td>Galc</td>
<td>12</td>
<td>36</td>
<td>98232394-98232429</td>
</tr>
<tr>
<td>Mylk4</td>
<td>13</td>
<td>13</td>
<td>32711300-32711312</td>
</tr>
<tr>
<td>Fzd5</td>
<td>1</td>
<td>22</td>
<td>64737712-64737722</td>
</tr>
<tr>
<td>Lrh1</td>
<td>1</td>
<td>6</td>
<td>136849713-136870578</td>
</tr>
<tr>
<td>Lamc2</td>
<td>1</td>
<td>923</td>
<td>153145967-153191429</td>
</tr>
<tr>
<td>Pofut1</td>
<td>2</td>
<td>188</td>
<td>153239719-153275253</td>
</tr>
<tr>
<td>Ache</td>
<td>5</td>
<td>19</td>
<td>137288432-137288449</td>
</tr>
<tr>
<td>Zfp282</td>
<td>6</td>
<td>1016</td>
<td>47876551-47913485</td>
</tr>
<tr>
<td>Bm8-Mar</td>
<td>6</td>
<td>4231</td>
<td>116342541-116360897</td>
</tr>
<tr>
<td>Setd1a</td>
<td>7</td>
<td>1834</td>
<td>127778726-127802950</td>
</tr>
<tr>
<td>Phkb</td>
<td>8</td>
<td>3197</td>
<td>85837045-86008353</td>
</tr>
<tr>
<td>Atax</td>
<td>X</td>
<td>7</td>
<td>105863283-105906922</td>
</tr>
</tbody>
</table>

Figure 2 KEGG pathway annotations of Lrh1 and its homology genes.
and proliferation. For example, a previous genome-wide association study has identified five SNPs on 1q32.1 associated with pancreatic cancer that mapped to Lrh1 gene and its up-stream regulatory region [33].

Conclusions

In conclusion, FTF combined with CYP7A might play an important role in Lrh1 regulated pancreas-specific transcriptional network. Furthermore, the SNPs analysis of Lrh1 and its homology genes provided the candidate mutant sites that might affect the Lrh1-related production and secretion of pancreatic fluid. These common susceptibility loci for Lrh1 and its homologous genes needed follow-up studies.

Highlights

1. Total 15683 DEGs were identified, of which 23 was Lrh1 homology genes (3 up-regulated and 20 down-regulated).
2. Fetoprotein TF was the only TF of Lrh1 identified based on TRANSFAC database and the promoter-binding factor of fetoprotein TF was CYP7A.
3. The SNP annotations of Lrh1 homologous genes showed that 92% of mutation sites were occurred in intron and upstream. Three SNPs of Lrh1 were located in intron, while 1819 SNPs of Phkb were located in intron and 1343 SNPs were located in upstream region.

Competing interest

The authors declare that they have no competing interests.

Authors’ contributions

MT, XM and CL participated in the design of this study, and they both performed the statistical analysis. RX, GYH and YZ carried out the study, together with LQ, collected important background information and drafted the manuscript. HL, XWP and ZS conceived of this study, and participated in the design and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This study was supported by grants from the National Natural Science Foundation of China (No. 81200320, 81300350), Shanghai Science and Technology Commission (No. 11JC1440200, Fund of Shanghai Health Bureau (No. 20114315) and Training Plan of Excellent Academic Researcher of Foundation of China (No. 81200320, 81300350), Shanghai Science and Technology Committee (No. 12410502). The authors would like to thank all the members of the Department of Medicine, Shanghai Jiao Tong University Affiliated Sixth People’s Hospital for their contributions to this study.

Received: 20 December 2013 Accepted: 27 March 2014

Published: 15 April 2014

References


---

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at www.biomedcentral.com/submit
Dear Authors,

During production of your paper, the following queries arose. Please respond to these by annotating your proofs with the necessary changes/additions. If you intend to annotate your proof electronically, please refer to the E-annotation guidelines. We recommend that you provide additional clarification of answers to queries by entering your answers on the query sheet, in addition to the text mark-up.

<table>
<thead>
<tr>
<th>Query No.</th>
<th>Query</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>URL: Please check that the following URLs are working. If not, please provide alternatives: <a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a></td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td>URL: Please check that the following URLs are working. If not, please provide alternatives: <a href="http://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=std">http://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=std</a></td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>URL: Please check that the following URLs are working. If not, please provide alternatives: <a href="http://www.nursa.org/molecule.cfm?molType=receptor&amp;molId=5A2">http://www.nursa.org/molecule.cfm?molType=receptor&amp;molId=5A2</a></td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td><a href="http://repositorio.unican.es/xmlui/bitstream/handle/10902/555/%5B2%5D%20Quiles%20Romagosa%20MA.pdf?sequence=1">http://repositorio.unican.es/xmlui/bitstream/handle/10902/555/%5B2%5D%20Quiles%20Romagosa%20MA.pdf?sequence=1</a></td>
<td></td>
</tr>
</tbody>
</table>