Identification of genes and signaling pathways associated with arthrogryposis-renal dysfunction-cholestasis syndrome using weighted correlation network analysis

MIAO CHAI*, LIJU SU*, XIAOLEI HAO, MENG ZHANG, LIHUI ZHENG, JIABING BI, XIAO HAN and BOHAI YU

Department of Clinical Laboratory, The First Hospital of Harbin, Harbin, Heilongjiang 150010, P.R. China

Received July 15, 2017; Accepted June 7, 2018

DOI: 10.3892/ijmm.2018.3768

Abstract. The present study aimed to identify the molecular basis of the arthrogryposis-renal dysfunction-cholestasis (ARC) syndrome, which is caused by mutations in the vacuolar protein sorting 33 homolog B (*VPS33B)* gene. The microarray dataset GSE83192, which contained six liver tissue samples from *VPS33B* knockout mice and four liver tissue samples from control mice, was downloaded from the Gene Expression Omnibus database. The differentially expressed genes (DEGs) were screened by the Limma package in R software. The DEGs most relevant to ARC were selected via weighted gene co-expression network analysis to construct a protein-protein interaction (PPI) network. In addition, module analysis was performed for the PPI network using the Molecular Complex Detection function. Functional and pathway enrichment analyses were also performed for DEGs in the PPI network. Potential drugs for ARC treatment were predicted using the Connectivity Map database. In total, 768 upregulated and 379 downregulated DEGs were detected in the *VPS33B* knockout mice, while three modules were identified from the PPI network constructed. The DEGs in module 1 (*CD83*, *IL1B* and *TLR2*) were mainly involved in the positive regulation of cytokine production and the Toll-like receptor (TLR) signaling pathway. The DEGs in module 2 (*COL1A1* and *COL1A2*) were significantly enriched with respect to cellular component organization, extracellular matrix-receptor interactions and focal adhesion. The DEGs in module 3 (*ABCG8* and *ABCG3*) were clearly associated with sterol absorption and transport. Furthermore, mercaptopurine was identified to be a potential drug (connectivity score=-0.939) for ARC treatment. In conclusion, the results of the current study may help to further understand the pathology of ARC, and the DEGs identified in these modules may serve as therapeutic targets.

Correspondence to: Dr Bohai Yu, Department of Clinical Laboratory, The First Hospital of Harbin, 151 Diduan Street, Daoli, Harbin, Heilongjiang 150010, P.R. China
E-mail: hongseredai@126.com

*Contributed equally

Key words: arthrogryposis-renal dysfunction-cholestasis, weighted gene co-expression network, module analysis, therapeutic drug

Introduction

Arthrogryposis-renal dysfunction-cholestasis (ARC) syndrome is a life-threatening autosomal recessive multisystem disorder caused by germline mutations in VPS33B-interacting protein, apical-basolateral polarity regulator (*VIPAR*) or vacuolar protein sorting 33 homolog B (*VPS33B*) (1). The principle clinical manifestations of ARC include renal tubular dysfunction, cholestasis, ichthyosis, central nervous system malformation and congenital joint contractures involving multiple organ systems (2,3). It has been recognized that ARC syndrome exhibits notable clinical variability, and the prognosis of this condition is particularly poor, with the majority of patients not surviving beyond the first year of life (4,5). Furthermore, there is currently no specific treatment for this syndrome.

Mutations in *VPS33B* are detectable in 75-77% of patients with a clinical diagnosis of ARC syndrome (3,6). A better understanding of the molecular pathology of this disorder is of vital importance for the development of an appropriate therapeutic regimen. *VPS33B* encodes a 617-amino-acid protein, which is a homolog of the class C yeast vacuolar protein sorting, and the VPS33B protein contains a Sec-1 domain involved in intracellular protein sorting and vesicular trafficking (7). It has also been reported that *VPS33B* is a downstream target gene of the *hnf6/vhnf1* signaling pathway that is important for zebrafish biliary development (8). In addition, the VPS33B protein can interact with soluble N-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs), which are involved in vesicular exocytosis and synaptic transmission to facilitate vesicle targeting and fusion (9). Therefore, the interaction between the mutant protein expressed in *VPS33B* mutants and the SNAREs at the late endosomal stage may be impeded, leading to abnormal secretion of lamellar granules, and localization or accumulation of plasma proteins (2,10). Abnormal protein trafficking and impairment in the maturation of multi-vesicular bodies in megakaryocytes underlie the α-granule deficiency in a mouse model of *VPS33B* deficiency and in patients with ARC (11). The VPS33B-VIPAR complex may regulate apical-basolateral polarity via the Rab11a-dependent apical recycling pathway and the transcriptional regulation of epithelial cadherin (1).

This complex also regulates the delivery of lysyl hydroxylase 3 into newly identified post-Golgi collagen IV carriers, which are essential for the modification of lysine residues in multiple collagen types (12).

In the study by Hanley *et al* (13), a murine model with a liver-specific deletion of *VPS33B* (*VPS33B* $^{fl/fl}$-AlfpCre) was successfully established, as indicated by the abnormalities identified in mice, which were similar to those observed in children with ARC syndrome. Furthermore, the analysis of gene expression profiles provided an insight into the possible regulatory mechanisms responsible for ARC syndrome. However, only the gene expression and pathway analysis of the microarray data were performed in the aforementioned study. To further elucidate the molecular basis of ARC, the gene expression profiles deposited by Hanley *et al* (13) were downloaded in the present study in order to conduct weighted gene co-expression network analysis and to identify potential therapeutic drugs.

Materials and methods

Microarray data and preprocessing. The gene expression profile of GSE83192 (13), generated by the GPL16570 platform (Affymetrix Mouse Gene 2.0 Array; Thermo Fisher Scientific, Inc., Waltham, MA, USA), was downloaded from Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/). This dataset contained six liver tissue samples from liver-specific *VPS33B* knockout (*VPS33B*$^{fl/fl}$-AlfpCre) mice and four liver tissue samples from control (*VPS33B*$^{fl/fl}$) mice. The raw data preprocessing was conducted using the oligo package in R software (www.r-project.org), including conversion of the data format, filling-in of missing data with the median values (14), background correction using the MicroArray Suite method (15) and normalization of the sequencing data using the quantile method (16).

Differential expression analysis and hierarchical clustering. The Limma package (17) was used to perform differential expression analysis for normalized values. In addition, P-values were adjusted for the false discovery rate (FDR) via the method described by Benjamini and Hochberg (18). The thresholds for differentially expressed gene (DEG) screening were set at FDR<0.05 and $|\log_2$ fold change$|>1$. The expression values of screened DEGs were hierarchically clustered by the pheatmap package (19) in R to intuitively observe the differences in gene expression levels.

Identification of co-expression modules. The weighted gene co-expression network analysis (WGCNA) package in R (20) was used to identify disease-associated co-expression modules. The default method defines the co-expression similarity S_{mn} as the absolute value of the correlation coefficient between the profiles of nodes m and n, namely $S_{mn} = |\text{cor}_{(m,n)}|$. The weighted adjacency was defined as $\alpha_{mn} = \text{power}_{(Smn,\beta),}$ and β was selected according to the scale-free topology criterion. The β represented the correlation coefficient between log (k) and log P (k), where k was the number of the nodes and P (k) was the probability of occurrence. The topological overlap measure between gene expression traits was calculated according to the formula $\varpi_{mn} = \frac{l_{mn} + \alpha_{mn}}{\min\{k_m, k_n\} + 1 - \alpha_{mn}}$ (21), where l_{mn} represented the number of nodes connected with both m and n, while k_m and k_n denoted the network connectivity of the nodes m and n, respectively. A topological overlap measure-based dissimilarity measure ($d_{mn} = 1-\omega_{mn}$) was used for dendrogram construction.

The gene significance (GS) values, defined as the log of the P-value, indicated the difference in the mRNA expression between *VPS33B* knockout and control mice. The module significance (MS), defined as the mean value of GS for all genes in a given module, was calculated for each module to identify its connection with the disease status. Two representative co-expression modules with the highest MS values were selected since a higher MS value indicates a closer connection.

Construction and analysis of the protein-protein interaction (PPI) network. The DEGs in the two selected representative co-expression modules mentioned earlier were adopted for PPI network construction. The database Search Tool for the Retrieval of Interacting Genes/Proteins (22) was employed to collect pairwise PPIs among the DEGs. Cytoscape software (23) was applied for visualization of the interaction associations, and the Molecular Complex Detection (MCODE) plugin (24) was used to create the modules with the following parameters: Degree cut-off=2, node score cut-off=0.2, and K-core=2. BiNGO (25), another plugin of Cytoscape, was used to annotate module function with an adjusted P-value of <0.05.

Enrichment analysis and potential therapeutic drug identification. The Gene Ontology (GO) annotations of the PPI network were performed by GOstat (26) in three categories, namely biological process (BP), cellular component (CC) and molecular function (MF), with P<0.05. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was conducted by the KEGG Orthology-Based Annotation System server (27), and P<0.05 was used as the cut-off criterion. Bioactive small molecules of putative relevance to the ARC syndrome were searched for using the Connectivity Map (CMAP) database, with the criteria set to |connectivity score|>0.9 and P<0.05 (28). A connectivity score closer to -1 implied that this small molecule may have a stronger therapeutic effect.

Results

DEG screening and hierarchical clustering. Subsequent to data preprocessing, 1,016 DEGs, including 768 upregulated and 248 downregulated genes, were identified in the *VPS33B* knockout mice as compared with the control mice. The expression values of screened DEGs were hierarchically clustered by pheatmap package, and the color contrast indicated that there were significant differences in gene expression between the *VPS33B*$^{fl/fl}$-AlfpCre and *VPS33B*$^{fl/fl}$ mice (Fig. 1).

WGCNA analysis and PPI network construction. Based on the correlation coefficient between log (k) and log P (k), the value of the adjacency matrix was set to 18 in order to guarantee the scale-free topology of the co-expression modules (Fig. 2). A gene clustering tree with the cut-off height of 0.9 was then constructed (Fig. 3), and the MS values of all modules were >0.6 with a P<0.05 (Fig. 4). Two representative co-expression modules, including the green module (72 DEGs) and the

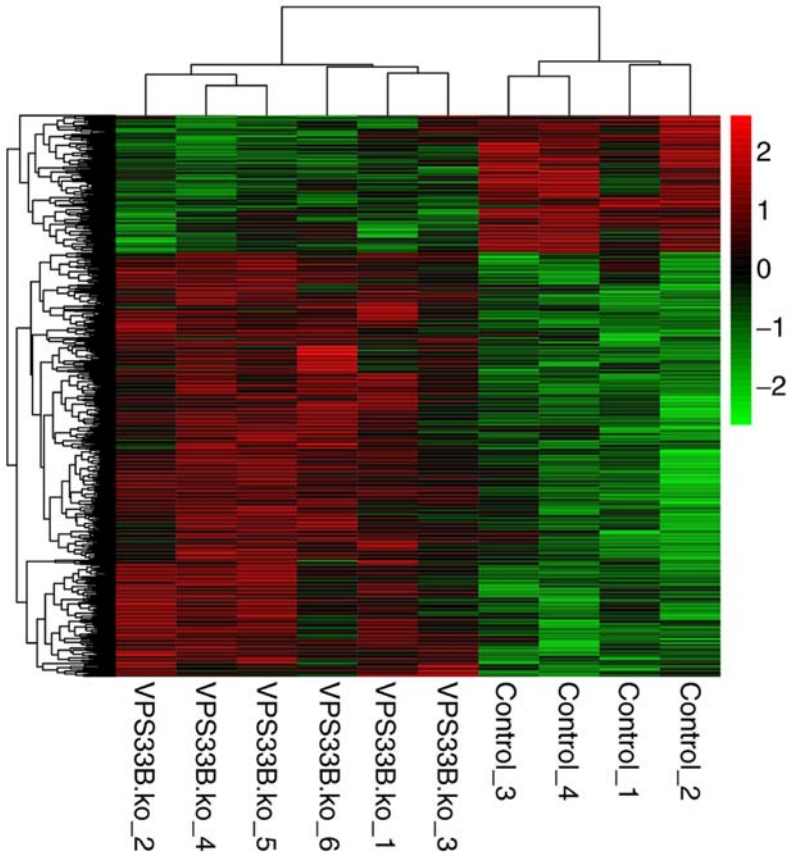

Figure 1. Differences in the expression values of differentially expressed genes between liver tissue samples from the *VPS33B* (*VPS33B*$^{fl/fl}$-AlfpCre) knockout mice and control (*VPS33B*$^{fl/fl}$) mice. Vps33b, vacuolar protein sorting 33 homolog B.

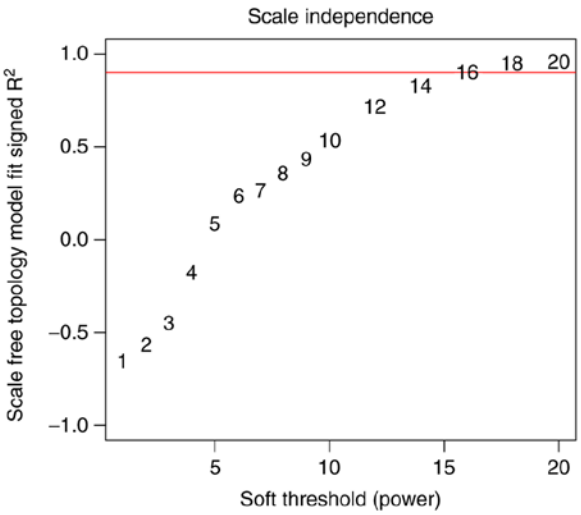

Figure 2. Power value of the adjacency matrix. The red line represents the correlation coefficient between log (k) and log P (k) of 0.9.

turquoise module (247 DEGs), were selected. Subsequently, the PPI network was constructed according to the PPIs of DEGs in these two representative co-expression modules, and the constructed PPI network contained 71 nodes and 135 PPIs (Fig. 5).

PPI network analysis. Three modules were identified from the constructed PPI network by MCODE (Fig. 5). The functional annotations revealed that the genes in module 1 were mainly involved in the positive regulation of biological processes (*CD86*, *CD83*, *IL1B* and *TLR2*; adjusted P=1.32x10^{-3}) and positive regulation of cytokine production (*CD83*, *IL1B* and *TLR2*; adjusted P=5.25x10^{-5}; Table I). The DEGs in module 2 were significantly enriched with respect to protein heterotrimerization (*COL1A1* and *COL1A2*; adjusted P=2.85x10^{-6}) and cellular component organization (*COL1A1*, *COL1A2* and *CD44*; adjusted P=1.12x10^{-2}). The four DEGs (*ABCG8*, *ABCG5*, *ABCB4* and *ABCG3*) in module 3 were clearly associated with transport (adjusted P=1.49x10^{-3}) and the establishment of localization (adjusted P=1.49x10^{-3}). *ABCG8* was observed to mainly participate in intestinal cholesterol absorption, lipid digestion, cholesterol efflux, intestinal absorption and sterol transport.

Functional and pathway enrichment analysis of the PPI network. The functional enrichment analysis revealed that the DEGs in the PPI network were significantly correlated with 10, 11 and 11 GO terms in the BP, CC and MF categories, respectively (Table II). A total of 11 DEGs (*ALCAM*, *ITGAX*, *CLDN4*, *CD44*, *ITGB8*, *CD34*, *CLDN6*, *BCL2*, *CDH1*, *CD2AP* and *SPP1*) were mainly involved in cell adhesion (P=7.25x10^{-5}). In addition, 7 DEGs (*ALCAM*, *CD83*, *CD86*, *ITGAX*, *CD44*, *CD34* and *TLR2*) were significantly associated with the external side of the plasma membrane (P=4.44x10^{-4}) and cell surface (P=4.44x10^{-3}). Meanwhile, the DEGs in the PPI network were evidently associated with ATP-binding transport activity (*ABCG8*, *ABCG5*, *ABCG3*, *ABCC5*, *ABCB4* and

Table I. Enriched GO terms for DEGs in the three identified modules from the protein-protein interaction network.

GO ID	Adjusted P-value	Description	DEGs
Module 1			
GO:0048518	1.32×10^{-3}	Positive regulation of biological process	*CD86*, *CD83*, *IL1B, TL*R2
GO:0050789	1.82×10^{-2}	Regulation of biological process	*CD86*, *CD83*, *IL1B*, *TLR2*
GO:0065007	2.22×10^{-2}	Biological regulation	*CD86*, *CD83*, *IL1B*, *TLR2*
GO:0001819	5.25×10^{-5}	Positive regulation of cytokine production	*CD83*, *IL1B*, *TLR2*
GO:0031349	5.25×10^{-5}	Positive regulation of defense response	*CD86*, *IL1B*, *TLR2*
GO:0031347	1.54×10^{-4}	Regulation of defense response	*CD86*, *IL1B*, *TLR2*
GO:0001817	1.54×10^{-4}	Regulation of cytokine production	*CD83*, *IL1B*, *TLR2*
GO:0048584	2.31×10^{-4}	Positive regulation of response to stimulus	*CD86*, *IL1B*, *TLR2*
GO:0002684	2.31×10^{-4}	Positive regulation of immune system process	*CD86*, *CD83*, *TLR2*
GO:0080134	2.31×10^{-4}	Regulation of response to stress	*CD86*, *IL1B*, *TLR2*
GO:0051240	2.31×10^{-4}	Positive regulation of multicellular organismal process	*CD83*, *IL1B*, *TLR2*
GO:0002682	4.86×10^{-4}	Regulation of immune system process	*CD86*, *CD83*, *TLR2*
GO:0048583	6.88×10^{-4}	Regulation of response to stimulus	*CD86*, *IL1B*, *TLR2*
GO:0010033	1.76×10^{-3}	Response to organic substance	*CD83*, *IL1B*, *TLR2*
GO:0002376	1.80×10^{-3}	Immune system process	*CD86*, *IL1B*, *TLR2*
GO:0051239	3.31×10^{-3}	Regulation of multicellular organismal process	*CD83*, *IL1B*, *TLR2*
GO:0042221	3.31×10^{-3}	Response to chemical stimulus	*CD83*, *IL1B*, *TLR2*
GO:0048519	5.00×10^{-3}	Negative regulation of biological process	*CD83*, *IL1B*, *TLR2*
GO:0048522	5.02×10^{-3}	Positive regulation of cellular process	*CD83*, *IL1B*, *TLR2*
GO:0050896	1.10×10^{-2}	Response to stimulus	*CD83*, *IL1B*, *TLR2*
Module 2			
GO:0070208	2.85×10^{-6}	Protein heterotrimerization	*COL1A1*, *COL1A2*
GO:0070206	3.99×10^{-5}	Protein trimerization	*COL1A1*, *COL1A2*
GO:0051291	2.91×10^{-3}	Protein heterooligomerization	*COL1A1*, *COL1A2*
GO:0051259	6.07×10^{-3}	Protein oligomerization	*COL1A1*, *COL1A2*
GO:0070271	9.81×10^{-3}	Protein complex biogenesis	*COL1A1*, *COL1A2*
GO:0006461	9.81×10^{-3}	Protein complex assembly	*COL1A1*, *COL1A2*
GO:0016043	1.12×10^{-2}	Cellular component organization	*COL1A1*, *COL1A2*, *CD44*
GO:0065003	1.12×10^{-2}	Macromolecular complex assembly	*COL1A1*, *COL1A2*
GO:0048646	1.12×10^{-2}	Anatomical structure formation involved in morphogenesis	*COL1A1*, *CD44*
GO:0043933	1.12×10^{-2}	Macromolecular complex subunit organization	*COL1A1*, *COL1A2*
Module 3			
GO:0006810	1.49×10^{-3}	Transport	*ABCG8*, *ABCG5*, *ABCB4*, *ABCG3*
GO:0051234	1.49×10^{-3}	Establishment of localization	*ABCG8*, *ABCG5*, *ABCB4*, *ABCG3*
GO:0051179	1.62×10^{-3}	Localization	*ABCG8*, *ABCG5*, *ABCB4*, *ABCG3*
GO:0030299	4.77×10^{-3}	Intestinal cholesterol absorption	*ABCG8*
GO:0044241	5.72×10^{-3}	Lipid digestion	*ABCG8*
GO:0033344	1.43×10^{-2}	Cholesterol efflux	*ABCG8*
GO:0050892	1.50×10^{-2}	Intestinal absorption	*ABCG8*
GO:0006855	1.67×10^{-2}	Drug transmembrane transport	*ABCB4*
GO:0015893	1.67×10^{-2}	Drug transport	*ABCB4*
GO:0015918	1.67×10^{-2}	Sterol transport	*ABCG8*

GO, gene ontology; DEGs, differentially expressed genes.

ABCA; P=4.71×10^{-3}), cytokine activity (*CCL2*, *CCL19*, *IL1B*, *CCL6* and *SPP1*; P=5.31×10^{-3}), and cholesterol transporter and sterol transporter activities (*ABCG8* and *ABCG5*; P=3.47×10^{-3}). Furthermore, *COL1A2* and *COL1A1* were associated with

Table II. Functional annotations in the BP, CC and MF categories for DEGs in the protein-protein interaction network.

Term	Description	P-value	Count	DEGs
BP				
GO:0007155	Cell adhesion	$7.25x10^{-5}$	11	*ALCAM*, *ITGAX*, *CLDN4*, *CD44*, *ITGB8*, *CD34*, *CLDN6*, *BCL2*, *CDH1*, *CD2AP*, *SPP1*
GO:0022610	Biological adhesion	$7.36x10^{-5}$	11	*ALCAM*, *ITGAX*, *CLDN4*, *CD44*, *ITGB8*, *CD34*, *CLDN6*, *BCL2*, *CDH1*, *CD2AP*, *SPP1*
GO:0006952	Defense response	$7.02x10^{-5}$	10	*NFKBIZ*, *TMEM173*, *CCL2*, *CD44*, *BCL2*, *TLR2*, *CCL19*, *IL1B*, *APAF1*, *TLR6*
GO:0009611	Response to wounding	$7.13x10^{-5}$	9	*NFKBIZ*, *CCL2*, *CD44*, *BCL2*, *TLR2*, *CCL19*, *IL1B*, *TLR6*, *PLAUR*
GO:0006955	Immune response	$5.71x10^{-4}$	9	*TMEM173*, *CCL2*, *TLR2*, *CCL19*, *IL1B*, *AF251705*, *TLR6*, *GBP3*, *CCL6*
GO:0006954	Inflammatory response	$2.84x10^{-4}$	7	*NFKBIZ*, *CCL2*, *CD44*, *TLR2*, *CCL19*, *IL1B*, *TLR6*
GO:0051240	Positive regulation of multicellular organismal process	$4.23x10^{-3}$	5	*CD83*, *CHRM3*, *BCL2*, *TLR2*, *IL1B*
GO:0006631	Fatty acid metabolic process	$6.48x10^{-3}$	5	*HACL1*, *CYP4A32*, *ACNAT2*, *ALOX5AP*, *ACACB*
GO:0016337	Cell-cell adhesion	$1.52\ x10^{-2}$	5	*CLDN4*, *CLDN6*, *BCL2*, *CDH1*, *CD2AP*
GO:0044093	Positive regulation of molecular function	$3.51x10^{-2}$	5	*CHRM3*, *BCL2*, *TLR2*, *APAF1*, *TLR6*
CC				
GO:0009897	External side of plasma membrane	$4.44x10^{-4}$	7	*ALCAM*, *CD83*, *CD86*, *ITGAX*, *CD44*, *CD34*, *TLR2*
GO:0045177	Apical part of cell	$4.78x10^{-4}$	6	*EPCAM*, *ABCG8*, *ABCG5*, *CDH1*, *ABCB4*, *SPP1*
GO:0009986	Cell surface	$3.32x10^{-3}$	7	*ALCAM*, *CD83*, *CD86*, *ITGAX*, *CD44*, *CD34*, *TLR2*
GO:0005584	Collagen type I	$9.57x10^{-3}$	2	*COL1A2*, *COL1A1*
GO:0016324	Apical plasma membrane	$1.20x10^{-2}$	4	*EPCAM*, *ABCG8*, *ABCG5*, *ABCB4*
GO:0000267	Cell fraction	$2.34x10^{-2}$	8	*ABCG5*, *CYP3A16*, *CYP4A32*, *CLIC5*, *BCL2*, *APAF1*, *ABCC5*, *ABCB4*
GO:0034707	Chloride channel complex	$2.39\ x10^{-2}$	3	*CLIC5*, *ANO1*, *ANO6*
GO:0030054	Cell junction	$2.47\ x10^{-2}$	7	*CHRM3*, *CLDN4*, *CLDN6*, *CDH1*, *PAK1*, *ABCB4*, *GRID1*
GO:0016323	Basolateral plasma membrane	$3.01x10^{-2}$	4	*EPCAM*, *CD44*, *CDH1*, *PAK1*
GO:0005583	Fibrillar collagen	$3.31\ x10^{-2}$	2	*COL1A2*, *COL1A1*
MF				
GO:0016887	ATPase activity	$4.71\ x10^{-3}$	6	*ABCG8*, *ABCG5*, *ABCG3*, *ABCC5*, *ABCB4*, *ABCA5*
GO:0005125	Cytokine activity	$0.53\ x10^{-3}$	5	*CCL2*, *CCL19*, *IL1B*, *CCL6*, *SPP1*
GO:0008009	Chemokine activity	$9.65x10^{-3}$	3	*CCL2*, *CCL19*, *CCL6*
GO:0042379	Chemokine receptor binding	$1.01\ x10^{-2}$	3	*CCL2*, *CCL19*, *CCL6*
GO:0005254	Chloride channel activity	$2.59x10^{-2}$	3	*CLIC5*, *ANO1*, *ANO6*
GO:0031404	Chloride ion binding	$2.90x10^{-2}$	3	*CLIC5*, *ANO1*, *ANO6*
GO:0005253	Anion channel activity	$2.98x10^{-2}$	3	*CLIC5*, *ANO1*, *ANO6*
GO:0043168g	Anion binding	$3.06x10^{-2}$	3	*CLIC5*, *ANO1*, *ANO6*
GO:0017127	Cholesterol transporter activity	$3.47\ x10^{-2}$	2	*ABCG8*, *ABCG5*
GO:0048407	Platelet-derived growth factor binding	$3.47\ x10^{-2}$	2	*COL1A2*, *COL1A1*
GO:0015248	Sterol transporter activity	$3.47\ x10^{-2}$	2	*ABCG8*, *ABCG5*

BP, biological process; CC, cellular component; MF, molecular function; DEGs, differentially expressed genes.

Table III. Significantly enriched pathways for DEGs in the protein-protein interaction network.

Term	Count	P-value	DEGs
mmu02010: ABC transporters	6	1.12×10^{-5}	*ABCG8*, *ABCG5*, *ABCG3*, *ABCC5*, *ABCB4*, *ABCA5*
mmu04514: Cell adhesion molecules	7	5.31×10^{-4}	*ALCAM*, *CD86*, *CLDN4*, *ITGB8*, *CD34*, *CLDN6*, *CDH1*
mmu04512: ECM-receptor interaction	5	2.28×10^{-3}	*CD44*, *ITGB8*, *COL1A2*, *COL1A1*, *SPP1*
mmu04620: Toll-like receptor signaling pathway	5	4.32×10^{-3}	*CD86*, *TLR2*, *IL1B*, *TLR6*, *SPP1*
mmu04510: Focal adhesion	6	1.03×10^{-2}	*ITGB8*, *BCL2*, *COL1A2*, *COL1A1*, *PAK1*, *SPP1*
mmu00640: Propanoate metabolism	3	1.74×10^{-2}	*ALDH1B1*, *ACACB*, *ACAT1*
mmu00620: Pyruvate metabolism	3	3.12×10^{-2}	*ALDH1B1*, *ACACB*, *ACAT1*
mmu00071: Fatty acid metabolism	3	3.71×10^{-2}	*CYP4A32*, *ALDH1B1*, *ACAT1*

DEGs, differentially expressed genes.

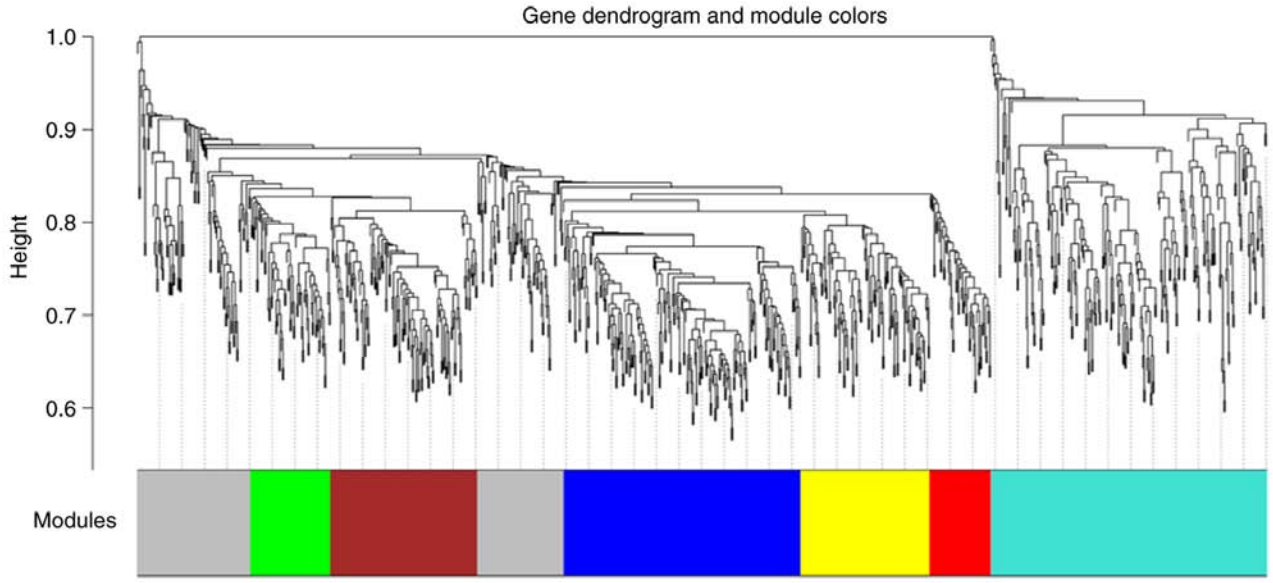

Figure 3. Gene dendrogram of co-expression modules.

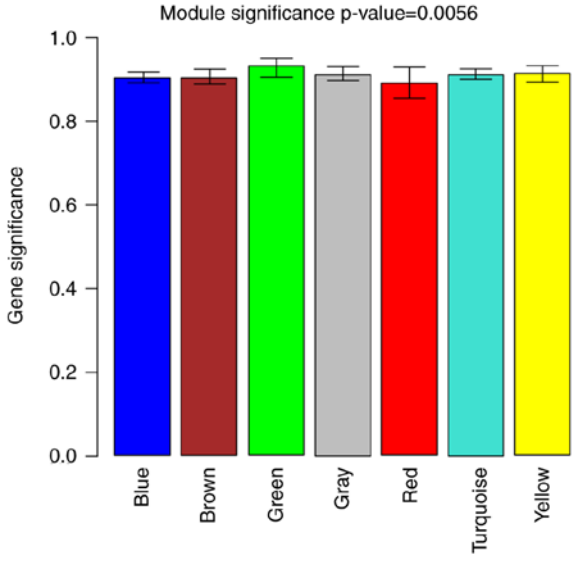

Figure 4. Module significance value of different co-expression modules.

collagen type I ($P=9.57\times10^{-3}$), fibrillar collagen ($P=3.31\times10^{-3}$) and platelet-derived growth factor binding ($P=3.47\times10^{-3}$).

In total, 8 KEGG pathways were identified for the DEGs in the PPI network (Table III). The members of the adenosine triphosphate-binding cassette (ABC) family, such as *ABCG8*, *ABCG5*, *ABCG3*, *ABCC5*, *ABCB4* and *ABCA5*, were significantly associated with the ABC transporters ($P=1.12\times10^{-5}$). Certain other DEGs, including *ITGB8*, *COL1A2*, *COL1A1* and *SPP1*, were evidently associated with extracellular matrix (ECM)-receptor interactions ($P=2.28\times10^{-3}$) and focal adhesion ($P=1.03\times10^{-2}$). Finally, the enriched Toll-like receptor (TLR) signaling pathway was associated with *CD86*, *TLR2*, *IL1B*, *TLR6* and *SPP1* ($P=4.31\times10^{-3}$).

Potentially therapeutic small molecules. A total of six small molecules, including mercaptopurine, ikarugamycin, camptothecin, quinostatin, dexpanthenol and DL-thiorphan, were screened using the CMAP database (Table IV). The score for mercaptopurine was the lowest (connectivity score=-0.939),

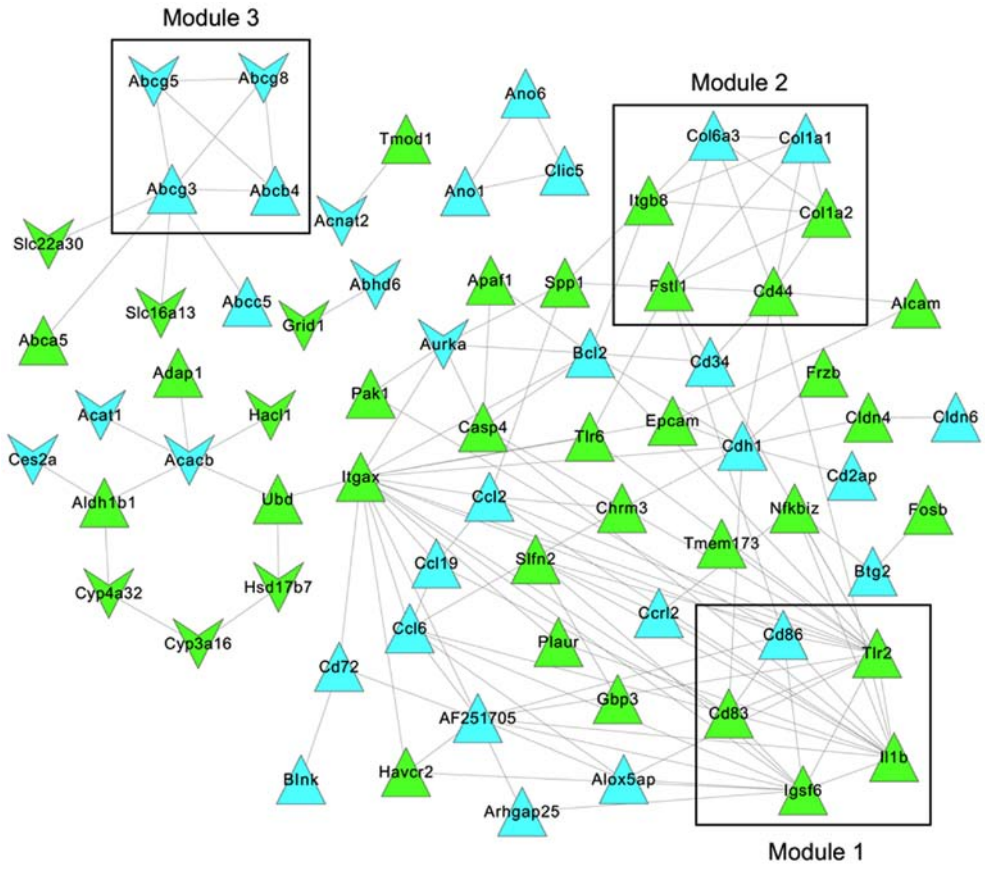

Figure 5. Protein-protein interaction network and the three identified modules. The regular triangle and inverted triangle stand for upregulated and down-regulated differentially expressed genes, respectively. The lines indicate the interactions between genes, while the colors of the triangles represent different co-expression modules.

Table IV. Identification of small molecules with a potential therapeutic role in arthrogryposis-renal dysfunction-cholestasis syndrome using the Connectivity Map database.

Name	Connectivity score	P-value
Mercaptopurine	-0.939	7.67×10^{-3}
Ikarugamycin	-0.906	1.62×10^{-3}
Camptothecin	0.902	1.82×10^{-3}
Quinostatin	0.904	1.88×10^{-2}
Dexpanthenol	0.920	4.00×10^{-5}
DL-thiorphan	0.975	9.70×10^{-4}

indicating that this small molecule may be a potential drug for ARC treatment.

Discussion

ARC, mainly caused by mutations in *VPS33B*, is associated with abnormalities in polarized liver and kidney cells, resulting in a multisystem disorder (1). In the present study, the microarray data of liver tissue samples from liver-specific *VPS33B* knockout mice and control mice were comprehensively analyzed. The DEGs in two representative co-expression modules with the highest MS values were selected for PPI network construction via WGCNA analysis. Three further modules were identified from the PPI network and annotated.

The five DEGs in module 1 included *CD86*, *CD83*, *IL1B*, *TLR2* and *LGSF6*. Pathway enrichment analysis of the PPI network demonstrated that *CD86*, *TLR2*, *IL1B*, *TLR6* and *SPP1* were significantly associated with the TLR signaling pathway (P=0.004317). The TLRs are part of the naive immune system and serve key roles in the elicitation of immune responses to microbes (29). It has been suggested that the VPS33B-VIPAR complex interacts with an active form of Rab11a (1). In addition, Rab11a-positive endosomes have been revealed to be important intermediates in the transport of TLRs (TLR2 and TLR4) and TLR adaptor molecules to phagosomes (30,31). In a study by Yu *et al* (32), deletion of *Rab11a* induced cytokine production and altered the intracellular distribution of TLRs, indicating that Rab11a contributes to intestinal host-microbial homeostasis through the sorting of TLRs. The data of the present study revealed that *CD83*, *IL1B* and *TLR2* were significantly enriched with respect to the positive regulation of cytokine production (adjusted $P=5.25 \times 10^{-5}$). Thus, the identified DEGs, including *CD83*, *IL1B*, *TLR2* and *TLR6*, may participate in the pathology of the ARC syndrome caused by mutations in *VPS33B* via the TLR signaling pathway and positive regulation of the cytokine production.

VPS33B serves a key role in the regulation of vesicle-to-target SNARE complex formation and subsequent membrane

fusion (33). Furthermore, inhibition of SNARE-mediated membrane traffic disrupted the intracellular integrin trafficking that can provide a linkage between the ECM and the cytoskeleton (34,35). In the present study, *COL1A2* and *COL1A1* in module 2 were significantly enriched with respect to cellular component organization, ECM-receptor interaction and focal adhesion. It has also been demonstrated that loss of *SNAP29* may cause alterations in the Rab11-expressing domains of the endocytic recycling compartment and the structure of focal adhesions, impairing endocytic recycling and cell motility (36). Taken together, the current study results provide evidence that mutations in *VPS33B* may disturb cellular component organization, ECM-receptor interactions and focal adhesion by regulating *COL1A2* and *COL1A1*.

It has been reported that vesicles containing ABC transporters co-localize with Rab11a prior to their insertion into the canalicular membrane (37). In the present study, the four DEGs (*ABCG8*, *ABCG5*, *ABCB4* and *ABCG3*) in module 3 were significantly involved in the ABC transporter pathway ($P=1.12x10^{-5}$). The ABC transporters are necessary for the energy-dependent biliary secretion of bile acids, phospholipids, sterols (for instance, ABCG8 and ABCG5 are sterol transporters) and non-bile acid organic anions (38). Impaired bile acid transport at the canalicular membrane, associated with reduced amounts of ABC transporter proteins, may cause cholestasis (bile secretory failure) (39). The functional annotations for DEGs in the PPI network revealed that *ABCG8* and *ABCG5* were evidently associated with cholesterol transporter and sterol transporter activities. In addition, *ABCG8* mainly participated in intestinal cholesterol absorption, lipid digestion, cholesterol efflux, intestinal absorption and sterol transport, according to the module annotations. Therefore, it may be speculated that mutations in *VPS33B* influence sterol absorption and transport by regulating *ABCG8* and *ABCG5*.

In conclusion, the results of the present study strongly indicate that the DEGs in the three identified modules serve important roles in the pathogenesis of ARC caused by mutations in *VPS33B*. Furthermore, *CD83*, *IL1B*, *TLR2* and *TLR6* may participate in the pathology by influencing the TLR signaling pathway and positive regulation of cytokine production. The mutations in *VPS33B* may disturb the cellular component organization, ECM-receptor interaction and focal adhesion by dysregulation of *COL1A2* and *COL1A*. Finally, sterol absorption and transport may also be impeded by mutations in *VPS33B* via the regulation of *ABCG8* and *ABCG5* expression.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

XHan, MZ and LZ searched and downloaded gene expression profile from the Gene Expression Omnibus database. MC, LS, JB, XHao and BY made substantial contributions to analysis and interpretation of microarray dataset. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Cullinane AR, Straatman-Iwanowska A, Zaucker A, Wakabayashi Y, Bruce CK, Luo G, Rahman F, Gürakan F, Utine E, Ozkan TB, *et al*: Mutations in VIPAR cause an arthrogryposis, renal dysfunction and cholestasis syndrome phenotype with defects in epithelial polarization. Nat Genet 42: 303-312, 2010.
2. Zhou Y and Zhang J: Arthrogryposis-renal dysfunction-cholestasis (ARC) syndrome: From molecular genetics to clinical features. Ital J Pediatr 40: 77, 2014.
3. Gissen P, Tee L, Johnson CA, Genin E, Caliebe A, Chitayat D, Clericuzio C, Denecke J, Di Rocco M, Fischler B, *et al*: Clinical and molecular genetic features of ARC syndrome. Hum Genet 120: 396-409, 2006.
4. Elmeery A, Lanka K and Cummings J: ARC syndrome in preterm baby. J Perinatol 33: 821-822, 2013.
5. Eastham KM, McKiernan PJ, Milford DV, Ramani P, Wyllie J, van't Hoff W, Lynch SA and Morris AA: ARC syndrome: An expanding range of phenotypes. Arch Dis Child 85: 415-420, 2001.
6. Cullinane AR, Straatman-Iwanowska A, Seo JK, Ko JS, Song KS, Gizewska M, Gruszfeld D, Gliwicz D, Tuysuz B, Erdemir G, *et al*: Molecular investigations to improve diagnostic accuracy in patients with ARC syndrome. Hum Mutat 30: E330-E337, 2009.
7. Carim L, Sumoy L, Andreu N, Estivill X and Escarceller M: Cloning, mapping and expression analysis of VPS33B, the human orthologue of rat *Vps33b*. Cytogenet Cell Genet 89: 92-95, 2000.
8. Matthews RP, Plumb-Rudewiez N, Lorent K, Gissen P, Johnson CA, Lemaigre F and Pack M: Zebrafish *vps33b*, an ortholog of the gene responsible for human arthrogryposis-renal dysfunction-cholestasis syndrome, regulates biliary development downstream of the onecut transcription factor *hnf6*. Development 132: 5295-5306, 2005.
9. Peterson M and Emr SD: The class C Vps complex functions at multiple stages of the vacuolar transport pathway. Traffic 2: 476-486, 2001.
10. Hershkovitz D, Mandel H, Ishida-Yamamoto A, Chefetz I, Hino B, Luder A, Indelman M, Bergman R and Sprecher E: Defective lamellar granule secretion in arthrogryposis, renal dysfunction, and cholestasis syndrome caused by a mutation in VPS33B. Arch Dermatol 144: 334-340, 2008.
11. Bem D, Smith H, Banushi B, Burden JJ, White IJ, Hanley J, Jeremiah N, Rieux-Laucat F, Bettels R, Ariceta G, *et al*: VPS33B regulates protein sorting into and maturation of α-granule progenitor organelles in mouse megakaryocytes. Blood 126: 133-143, 2015.
12. Banushi B, Forneris F, Straatman-Iwanowska A, Strange A, Lyne AM, Rogerson C, Burden JJ, Heywood WE, Hanley J, Doykov I, *et al*: Regulation of post-Golgi LH3 trafficking is essential for collagen homeostasis. Nature Communications 7: 12111, 2016.

13. Hanley J, Dhar DK, Mazzacuva F, Fiadeiro R, Burden JJ, Lyne AM, Smith H, Straatman-Iwanowska A, Banushi B, Virasami A, *et al*: Vps33b is crucial for structural and functional hepatocyte polarity. J Hepatol 66: 1001-1011, 2017.
14. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D and Altman RB: Missing value estimation methods for DNA microarrays. Bioinformatics 17: 520-525, 2001.
15. Hubbell E, Liu WM and Mei R: Robust estimators for expression analysis. Bioinformatics 18: 1585-1592, 2002.
16. Rao Y, Lee Y, Jarjoura D, Ruppert AS, Liu CG, Hsu JC and Hagan JP: A comparison of normalization techniques for microRNA microarray data. Stat Appl Genet Mol Biol 7: Article22, 2008.
17. Smyth GK: Limma: Linear models for microarray data. In: Bioinformatics and computational biology solutions using R and Bioconductor Springer, pp397-420, 2005.
18. Benjamini Y, and Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Statist Soc B 57: 289-300, 1995.
19. Wang L, Cao C, Ma Q, Zeng Q, Wang H, Cheng Z, Zhu G, Qi J, Ma H, Nian H and Wang Y: RNA-seq analyses of multiple meristems of soybean: Novel and alternative transcripts, evolutionary and functional implications. BMC Plant Biol 14: 169, 2014.
20. Langfelder P and Horvath S: WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics 9: 559, 2008.
21. Ravasz E, Somera AL, Mongru DA, Oltvai ZN and Barabási AL: Hierarchical organization of modularity in metabolic networks. Science 297: 1551-1555, 2002.
22. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C and Jensen LJ: STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41 (Database Issue): D808-D815, 2013.
23. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T: Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498-2504, 2003.
24. Bader GD and Hogue CW: An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4: 2, 2003.
25. Maere S, Heymans K and Kuiper M: BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21: 3448-3449, 2005.
26. Beissbarth T and Speed TP: GOstat: Find statistically over-represented Gene Ontologies within a group of genes. Bioinformatics 20: 1464-1465, 2004.
27. Wu J, Mao X, Cai T, Luo J and Wei L: KOBAS server: A web-based platform for automated annotation and pathway identification. Nucleic Acids Res 34 (Web Server Issue): W720-W724, 2006.
28. Cheng J, Yang L, Kumar V and Agarwal P: Systematic evaluation of connectivity map for disease indications. Genome Med 6: 540, 2014.
29. Beutler B: Inferences, questions and possibilities in Toll-like receptor signalling. Nature 430: 257-263, 2004.
30. Husebye H, Aune MH, Stenvik J, Samstad E, Skjeldal F, Halaas O, Nilsen NJ, Stenmark H, Latz E, Lien E, *et al*: The Rab11a GTPase controls Toll-like receptor 4-induced activation of interferon regulatory factor-3 on phagosomes. Immunity 33: 583-596, 2010.
31. Sjoelund V, Smelkinson M and Nita-Lazar A: Phosphoproteome profiling of the macrophage response to different toll-like receptor ligands identifies differences in global phosphorylation dynamics. J Proteome Res 13: 5185-5197, 2014.
32. Yu S, Nie Y, Knowles B, Sakamori R, Stypulkowski E, Patel C, Das S, Douard V, Ferraris RP, Bonder EM, *et al*: TLR sorting by Rab11 endosomes maintains intestinal epithelial-microbial homeostasis. EMBO J 33: 1882-1895, 2014.
33. Gissen P, Johnson CA, Morgan NV, Stapelbroek JM, Forshew T, Cooper WN, McKiernan PJ, Klomp LW, Morris AA, Wraith JE, *et al*: Mutations in VPS33B, encoding a regulator of SNARE-dependent membrane fusion, cause arthrogryposis-renal dysfunction-cholestasis (ARC) syndrome. Nat Genet 36: 400-404, 2004.
34. Skalski M and Coppolino MG: SNARE-mediated trafficking of alpha5beta1 integrin is required for spreading in CHO cells. Biochem Biophys Res Commun 335: 1199-1210, 2005.
35. Tayeb MA, Skalski M, Cha MC, Kean MJ, Scaife M and Coppolino MG: Inhibition of SNARE-mediated membrane traffic impairs cell migration. Exp Cell Res 305: 63-73, 2005.
36. Rapaport D, Lugassy Y, Sprecher E and Horowitz M: Loss of SNAP29 impairs endocytic recycling and cell motility. PLoS One 5: e9759, 2010.
37. Wakabayashi Y, Dutt P, Lippincott-Schwartz J and Arias IM: Rab11a and myosin Vb are required for bile canalicular formation in WIF-B9 cells. Proc Natl Acad Sci USA 102: 15087-15092, 2005.
38. Berge KE, Tian H, Graf GA, Yu L, Grishin NV, Schultz J, Kwiterovich P, Shan B, Barnes R and Hobbs HH: Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. Science 290: 1771-1775, 2000.
39. Strautnieks SS, Bull LN, Knisely AS, Kocoshis SA, Dahl N, Arnell H, Sokal E, Dahan K, Childs S, Ling V, *et al*: A gene encoding a liver-specific ABC transporter is mutated in progressive familial intrahepatic cholestasis. Nature genetics 20: 233-238, 1998.