Identification Of A Novel Disorder With Cholestasis, Liver Fibrosis And Congenital Diarrhea Due To Organic Solute Transporter Alpha Deficiency

Emily Gao

Follow this and additional works at: https://elischolar.library.yale.edu/ymltdl

Part of the Medicine and Health Sciences Commons

Recommended Citation

This Open Access Thesis is brought to you for free and open access by the School of Medicine at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Yale Medicine Thesis Digital Library by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.
Identification of a novel disorder with cholestasis, liver fibrosis and congenital diarrhea due to Organic Solute Transporter Alpha deficiency

A Thesis Submitted to the Yale University School of Medicine in Partial Fulfillment of the Requirements for the Degree of Doctor of Medicine

Emily Gao
Research Supervisor: Dr. Silvia Vilarinho
2020
ABSTRACT:

BACKGROUND: Much of our understanding of normal liver pathophysiology comes from studying patients with liver disease who remain undiagnosed after a comprehensive workup. Searching for diagnoses among such cohorts has benefits both for the underserved patients with idiopathic liver disease, but also for science as a whole. Unbiased genomic analysis of patients with undiagnosed disease has yielded remarkable results, including clinically actionable diagnoses in children and adults, as well as the discovery of novel genetic diseases. Thus, we posit that among pediatric patients with undiagnosed liver disease, there may be previously unrecognized Mendelian causes.

METHODS: We report a 2.5 year old with unexplained congenital chronic diarrhea, coagulopathy, fat-soluble vitamin deficiency, cholestasis and liver fibrosis. Exome sequencing and analysis was performed. Once a gene of interest was revealed, its expression at the mRNA level was assessed in colon tissue biopsy of the proband.

RESULTS: Exome sequencing uncovered a previously unidentified homozygous premature termination mutation (p.Gln186*) in SLC51A, encoding for the alpha subunit of the heteromeric organic solute transporter alpha-beta (OSTα-β), an important contributor to intestinal bile acid reabsorption in the enterohepatic circulation. Gene expression assay showed absence of SLC51A and detectable SLC51B mRNA transcripts in the proband’s colon tissue.

CONCLUSION: This study identifies a novel genetic digestive disorder due to OSTα deficiency. Homozygous loss-of-function mutation in SLC51A causes
chronic diarrhea, features of cholestasis, and liver fibrosis. Our findings further contribute to the body of evidence showing the importance of studying patients with undiagnosed disease, and the utility of whole exome sequencing as a powerful tool in clinical diagnosis and scientific discovery.
Published in part:


Please note that portions of the methods, results and discussion of this thesis were previously published in the above article, for which I was first author.

Presented in part:

# TABLE OF CONTENTS

ABSTRACT ........................................................................................................... ii

ACKNOWLEDGEMENTS ..................................................................................... 1

INTRODUCTION .................................................................................................. 2
  Role of human genetics in healthy and diseased liver ...................................... 3
  The enterohepatic circulation ........................................................................... 5
  Organic solute transporter α-β ......................................................................... 7
  Populations at greatest risk of genetic disease and highest potential for gene discoveries ............................................................... 8
  Whole Exome Sequencing .............................................................................. 9

STATEMENT OF PURPOSE .................................................................................. 12

PATIENTS AND METHODS .................................................................................. 13
  Human subjects
  DNA isolation, exome capture and sequencing .............................................. 13
  Exome sequencing analysis ........................................................................... 13
  Principal component analysis ....................................................................... 14
  Inbreeding coefficient .................................................................................... 15
  Sanger sequencing ......................................................................................... 15
  Bile acid measurement ................................................................................... 16

STATEMENT OF WORK ....................................................................................... 17

RESULTS ............................................................................................................... 18
  Case report ...................................................................................................... 18
  Homozygous loss-of-function mutation in SLC51A ...................................... 24
  SLC51A, but not SLC51B is absent in the proband's colon .............................. 30

DISCUSSION ......................................................................................................... 32

REFERENCES ...................................................................................................... 38
TABLE OF FIGURES

Figure 1. Schematic representation of the rising number of liver disease-causing genes discovered over the past 35 years......................................................... 4

Figure 2. Schematic representation of enterohepatic bile acid transporters...... 6

Figure 3. Overview of whole-exome sequencing pipeline ................................10

Figure 4. Liver histology of proband with undiagnosed cholestatic disease....... 22

Figure 5. Duodenal and colonic sections of proband........................................ 23

Figure 6. Pedigree and Sanger sequencing of the proband and his unaffected family members................................................................. 25

Figure 7. Inferring proband's ancestry .............................................................. 28

Figure 8. Schematic depiction of the location of the proband's terminating mutation................................................................................................. 29

Figure 9. Gene expression of SLC51A and SLC51B in proband and control colon tissues................................................................. 30
TABLE OF TABLES

Table 1. Clinical and laboratory findings in proband who harbors a homozygous premature termination mutation (P.Gln186*) in SLC51A………………………………………20

Table 2. Summary of laboratory tests performed as part of the comprehensive evaluation for abnormal liver function prior to pursuit of whole-exome sequencing………………………………………………………………………………………………………21

Table 3. Quality metrics of proband whole exome sequencing data…………………25

Table 4. Rare homozygous protein-altering variants identified in the proband…27
ACKNOWLEDGEMENTS

The work presented in this thesis is due to the mentorship of Dr. Silvia Vilarinho, who has patiently led and guided me through this scientific project, and role-modeled how to approach scientific challenges in the clinic and laboratory. Her vision of utilizing genetics to advance patient care and our understanding of liver disease has been inspiring. I would also like to thank Dr. Jungmin Choi for his guidance in identity by descent analysis. I would also like to express my gratitude to the contributions of the proband and his family as well as our collaborators.

Thank you to my parents, grandparents, brother and Dan for their tireless love and support.

This work was supported by the US National Institutes of Health (NIH) Centers for Mendelian Genomics (U54 HG006504) and NIH - Yale Liver Center P30DK034989.
INTRODUCTION

Despite major advances in the diagnosis and treatment of viral causes of hepatitis, the burden of chronic liver disease is rising worldwide, with end-stage liver disease ranking as the 11th leading cause of death worldwide in 2016 (1-3). As of 2018, over 4.5 million adults in the U.S. are diagnosed with chronic liver disease (4). The number of people affected by these diseases underscores the importance and urgency of understanding more about the pathophysiology of liver disease.

The liver is a complex, multi-purpose organ that serves as a hub of anabolism, catabolism and endocrine signaling, and its role in each of these areas remains incompletely understood. Our current knowledge gaps are reflected not only in our lack of treatments for metabolic diseases such as non-alcoholic fatty liver disease (NAFLD), but also in the number of patients with liver disease of unknown cause. Amongst various cohorts of patients with cirrhosis, 5-30% are labeled as “cryptogenic,” meaning that after exhaustive work-up, there is no known cause of their end-stage liver disease (5).

The prevalence and burden of idiopathic liver dysfunction is underestimated based on cryptogenic cirrhosis data alone, as it may remain asymptomatic until the first decompensation arises, such as ascites, esophageal variceal hemorrhage and/or hepatic encephalopathy. Elevated transaminases of unknown etiology have been reported in up to 2.5% of the general adult
A study of 254 adults with 6 months of persistently abnormal liver function tests identified no etiology for such abnormalities in nearly 20% (7). These studies demonstrate the gaps in knowledge in our understanding of the pathogenesis of liver disease, highlighting the importance of basic science and translational research to investigate the mechanisms of liver disease.

**Role of human genetics in healthy and diseased liver**

Historically, genetic studies of patients with liver disease have played a large role in our understanding of liver physiology (8). In the 1990s, genetics provided us with a molecular-level understanding of copper and iron-metabolism from studying patients with Wilson Disease and hemochromatosis, respectively (8, 9). Studying patients with Dubin-Johnson and Rotor Syndrome further elucidated the transporters involved in the flow of bile from the liver into the biliary tree and digestive tract (10, 11). By applying unbiased genomic analysis to undiagnosed pediatric liver disease, the Vilarinho laboratory has discovered recessive mutations in *DGUOK, ACOX2*, and *KIF12* as novel Mendelian causes of non-cirrhotic portal hypertension (12), bile acid synthesis disorder (13), and cholestatic liver disease (14), respectively. Within the last 35 years, nearly 100 cirrhosis-causing genes have been discovered and added to the Online Mendelian Inheritance in Man (OMIM) database (Figure 1).
Figure 1. Overview of the rising number of liver disease-causing genes discovered over the past 35 years. The graph depicts the approximate number of liver disease-causing genes identified within the last three decades based on the Online Mendelian Inheritance in Man (OMIM) database descriptive entries (last accessed in August 31, 2019). Specifically, we analyzed results from text query “cirrhosis” in the OMIM database. These genes were further investigated in Pubmed to find the original discovery date of the gene and liver-related phenotype. Somatic mutations, gene associations, and cases in which there was limited evidence of liver disease-gene causality were not included in this graph.

Despite major advances in human genetics since the completion of the human genome project nearly 20 years ago, the majority of genes still have partially or completely unknown function in normal human physiology. The putative functions of the majority of genes have been extrapolated from orthologous genes in model organisms such as yeast and mice (15). However, there are limits to understanding human physiology from orthologous organisms. For many genes there are significant differences in expression between humans and organisms such as mice, limiting our ability to draw meaningful conclusions about human disease from such experimental models. Moreover, many human gene
functions have been predicted through computational, in silico inferences rather than functional studies in model organisms (16).

Studying human phenotypes of monogenic disorders is a powerful tool for understanding a gene’s function. According to OMIM, mutations in approximately 20% of human protein-encoding genes have been shown to cause pathologic phenotypes, shedding light on their role in human health and disease (17). Given the liver’s multipurpose role in the body, understanding the pathophysiology of the liver has broadly generalizable medical benefits.

The enterohepatic circulation
All orally ingested drugs must pass through the liver before entering systemic circulation, and over half of the most widely prescribed drugs in the U.S. are metabolized predominantly in the liver (18). Understanding enterohepatic circulation on a molecular level is important to optimize medication efficacy and mitigate toxicity (19). Enterohepatic circulation refers to the chemical and physical recycling of bile. Bile is synthesized in the liver, excreted into the biliary ducts and released into the small intestine upon ingestion of a meal, after which it is reabsorbed by enterocytes and transported back to the liver through the portal vein. This circulation of bile enables the fat-soluble nutrients and pharmacologic agents from the alimentary track to be absorbed in the small intestine and colon and introduced to the portal vein and first-pass metabolism in the liver. As one of the body’s major sites for xenobiotic metabolism, the liver also secretes several
toxins, including mercury (20), as well as pharmacological agents, such as mycophenolate glucuronide, into bile, such that they can be excreted in feces or metabolized and reabsorbed by the body (21). Each of these steps is mediated on a molecular level by transporters, whose known locations are summarized in Figure 2.

Figure 2. Schematic representation of enterohepatic circulation highlighting major known bile acid transporters. Bile acids are depicted by yellow circles labeled “BA”. Because the OSTα-β transport of bile acids is bidirectional and diffusion-dependent, bidirectional arrows of different sizes are used to depict the flow of bile acids, with larger arrows indicating concentration-dependent flow in individuals without cholestasis. Image adapted from Ballatori et. al. 2009, *Frontiers in bioscience* (22).
Organic solute transporter α-β

Among the enterohepatic bile acid transporters, the heterodimer organic solute transporter α and organic solute transporter β (OSTα-β) serves as an example of a transporter that has been well-characterized in model organisms, with emerging relevance to human disease. OSTα-β is a heteromeric organic solute and steroid transporter that was initially discovered in 2001 in the liver of a marine vertebrate, *Leucoraja erinacea* (the little skate) (23). Two years later the human orthologs of these proteins, encoded by *SLC51A* and *SLC51B* were described (24). The former encodes for a 340–amino acid protein with seven transmembrane domains whereas the latter encodes for a 128–amino acid protein with a single transmembrane domain (24). Expression of both subunits is essential for transporter stability and delivery to the basolateral membrane of a variety of epithelial cells, and loss of expression of either subunit results in loss of transporter activity (23).

OSTα-β and the apical sodium-dependent bile acid transporter (ASBT) play important roles in the enterohepatic bile acid circulation. ASBT mediates sodium-dependent bile acid uptake across the luminal membrane of the enterocyte while transport across the basolateral membrane to the portal vein is facilitated by OSTα-β (25). In humans, OSTα-β is also expressed in the basolateral membrane of hepatocytes, cholangiocytes, colonocytes and renal proximal tubule cells, as well as testes, ovaries and adrenal glands (22, 24, 25). While recessive mutations in *SLC10A2*, which encodes ASBT, have been reported to cause a
human primary bile acid malabsorption disorder (26), and OSTβ deficiency has been identified in two siblings with chronic diarrhea and features of cholestasis with SLC51B mutations (27), no human disease has yet been ascribed to mutations in SLC51A. However, OSTα-β has recently received interest from the pharmaceutical industry, as its inhibition has been proposed as a potential mechanism of drug-induced liver injury (28). Moreover, OSTα-β has been implicated in playing a protective role in chronic cholestatic liver disease, as hepatic expression of these proteins are strongly up-regulated in patients with cholestatic liver injury (29).

**Populations at greatest risk of genetic disease and highest potential for gene discoveries**

Historically, studying patients with undiagnosed diseases among populations with a high rate of autozygosity, has led to the discovery of gene functions and novel genetic diseases, including Tay-Sachs in the Ashkenazi Jewish population or Progressive Familial Intrahepatic Cholestasis Type 1 in the Old Order Amish (30, 31), among many others. Children born of consanguineous unions are at higher risk of having recessive genetic diseases due to higher rates of autozygosity (32), with some studies estimating first-cousin marriage to have doubled risk compared to the general population (33). Other causes of high autozygosity include geographic isolation and endogamy (marriage within an ethnic or religious group). One of the countries with the highest rates of consanguinity is Pakistan. Given that up to 62% of marriages are consanguineous in certain
regions of the country (34), there is predicted to be a high burden of genetic disease.

**Whole Exome Sequencing**

Whole Exome Sequencing (WES) has emerged as a powerful and cost-effective tool in gene discovery and clinical diagnosis (35-37). WES captures and sequences protein-coding segments of DNA, known individually as exons, and in aggregate as the exome, which is predicted to encode for ~20,000 proteins. Although the exome makes up 1-2% of the human genome, variants in exons are suspected to account for 85% of genetic diseases (35). WES is comprised of two components, a wet laboratory process, and a computational pipeline, summarized in Figure 3. The wet laboratory process of WES begins with extraction of genomic DNA from biological samples (e.g., blood, buccal swabs, saliva, paraffin-embedded tissue, etc.), followed by selective capture and amplification of coding DNA (exons), and its massive parallel sequencing. The computational component of the WES pipeline transforms the raw sequence data into a dataset that can be analyzed for disease-causing genetic variants. WES data is aligned to and compared with the human reference genome, calling any variation from the reference sequence. In a process known as “annotation”, each variant’s minor allele frequency, chromosome and genome position, resulting amino acid change, predicted deleteriousness based on *in silico* calculations, and previous reports of pathogenicity are determined. Annotation data is then used to prioritize variants as disease-causing candidates based largely on the rareness and (predicted or known) pathogenicity of the variant.
WES has a reported diagnostic yield of 11-34% among cohorts of undiagnosed patients who had previously exhausted extensive workups (42-45). These diagnoses are significant not only to the patients suffering from these rare diseases, but often to the scientific community as a whole. When analyzing 160 patients through the NIH Undiagnosed Disease Network, novel associations were discovered between a gene and a known disease in 3.5% of patients, and novel genetic diseases were characterized in 1% of patients (45). Since 2014, 200-300 novel genetic diseases have been added to OMIM annually, with nearly 1000 novel genotype-phenotype relationships reported from 2016 up to now (46).
Over the past 10 years mutations in numerous genes have been discovered to cause cholestasis and other liver phenotypes, including genes with exclusive or predominant liver involvement and those with hepatic and extra-hepatic syndromes (13, 14, 47, 48). Here, by combining deep phenotyping with unbiased WES to study patient(s) suffering from liver disease of unknown etiology, we identified the first case of OSTα deficiency in a child with elevated liver transaminases, cholestasis and congenital diarrhea of unknown etiology.
STATEMENT OF PURPOSE

To identify novel genetic liver disorder(s) by combining deep phenotyping with unbiased WES in evaluation of patient(s) with unexplained liver disease.
PATIENTS AND METHODS

Human subjects
The proband and his family members were recruited for study. Written informed consent was provided for all participants. This study was approved by the institutional review boards of The Children’s Hospital and the Institute of Child Health in Lahore, Pakistan, and the Yale University.

DNA isolation, exome capture and sequencing
Genomic DNA was isolated from dried blood spots using standard procedures; the xGen capture kit from IDT was used for exome capture, and 100 base paired-end sequencing was performed on the Illumina HiSeq platform.

Exome sequencing analysis
WES data was aligned to the human reference genome (hg19) using BWA-MEM (38). Genome analysis tool kit (GATK) was used to identify nucleotide bases in the proband that differ from the reference genome, in a process known as “variant calling” (39). Variants with minor allele frequency of 0.01 or less were annotated with amino acid change consequences, predictions of deleteriousness, and minor allele frequencies utilizing Annovar (40). Allele frequencies were determined using the Genome Aggregation Database (gnomAD) which contains 125,748 exomes and 15,708 genomes from a diverse array of disease-specific studies as well as unrelated individuals sequenced for population genetics.
studies (last accessed on January 28, 2020). Rare mutations were subsequently prioritized by predicted deleteriousness. Prioritized mutations include nonsense, frameshift, start-lost or gained, splice site variants, and deleterious missense variants. Functional impact of protein-altering missense variants was predicted utilizing MetaSVM or Polyphen-2 (49), such that variants predicted to be benign were filtered out. Both MetaSVM and Polyphen-2 predictions are computed based on many factors, including sequence conservation among orthologous organisms, and protein structure modeling (49, 50). MetaSVM has been known to weigh conservation more heavily, based on the assumption that highly conserved regions are essential to healthy human physiology while poorly conserved regions may be less essential to protein structure and function (50). The alignment and predicted result of these variants of interest were manually confirmed utilizing BLAT, an alignment script embedded within the UCSC Human Genome Browser (51).

**Principal Component Analysis**

Ancestral origin was inferred using principal component analysis (PCA) of tagged single-nucleotide polymorphisms (SNPs). Tag SNP genotypes represent a region of the genome within a linkage disequilibrium block and are likely to be shared by a given subpopulation. Using EIGENSTRAT software, we performed PCA on tag SNPs extracted from our proband’s exome sequence combined with tag SNPs from HapMap project subjects of various ethnicities to infer the ancestry of our proband (52).
**Inbreeding coefficient**

Inbreeding coefficients were estimated by identifying homozygous by descent (HBD) using the software BEAGLE 3.2, as described previously (53).

**Sanger sequencing**

The nomenclature of the *SLC51A* variant is based on the National Center for Biotechnology Information reference sequence NM_152672. Sanger sequencing of the identified *SLC51A* mutation p.Gln186* was performed by PCR amplification of genome DNA of the proband, his parents and his sibling using the following primers: forward 5'-AGTGGGAGACCAAGAGGGTT-3', reverse 5'-CCTTACTCCCGGCTTACGTC-3'.

**mRNA gene expression TaqMan assays**

RNA was extracted from formalin fixed paraffin-embedded (FFPE) samples of proband’s colonic biopsies and unrelated normal-appearing colonic tissues using Beckmann Coulter agencourt FormaPure RNA kit (Ref#C19158). The superscript IV VILO kit was used for cDNA synthesis. Gene expression of *GAPDH* (Hs99999905_m1), *SLC51A* (Hs00380895_m1), and *SLC51B* (Hs01057182_m1) was assessed using ABI TaqMan assays.
**Bile acid measurement**

Bile acid concentrations in dried blood spots were measured using liquid chromatography – tandem mass spectrometry (LC-MS/MS) with stable-isotope internal standards as described previously (13, 54).
STATEMENT OF WORK

Dr. Vilarinho conceptualized and designed the project and under her guidance I performed the computational analysis for the proband's exome (including exome sequencing analysis, principal component analysis and identity by descent analysis), Taqman gene expression analysis, reviewed and integrated our results with the current literature and wrote this thesis.
RESULTS

Case Report

A two and a half year-old Pakistani male, born from a first cousin union, was found to have abnormal liver function tests (Table 1). He was born at full term, with no complications and normal weight, height and head circumference. Since birth, he has suffered from chronic malabsorptive diarrhea of unclear etiology complicated by failure to thrive. Of note, the patient was exclusively breast fed during his first year of life, with introduction of food only at 1 year of age. His past medical history is also significant for easy bruising, with two episodes of prolonged bleeding that required blood transfusions (one at the circumcision site at 1.5 months of age, and another at a vaccination site at 6 months of age), and two hospitalizations for respiratory infections of unclear etiology. Developmental milestones include social smile at 2 months, delayed sitting without support at 10 months of age and walking at 18 months of age attributed to motor delay in the setting of malnutrition due to underlying disease. At one year of age, he was started on nutritional supplementation, including a soy-based formula and fat-soluble vitamin (A, D, E, K), after which his weight and height improved significantly. However, at age two and a half he was found to have elevated aminotransferases and alkaline phosphatase (Table 1). A comprehensive evaluation was unrevealing of the etiology of liver dysfunction (Table 2). An abdominal ultrasound showed a mildly enlarged liver with normal liver echotexture and a mildly enlarged spleen spanning 8.6 cm. Liver biopsy at 3.8 years old revealed a lobular architecture suggestive of early cirrhosis with portal
and peri-portal fibrosis and many thin fibrous septa (Figure 4A,B). Portal areas showed only minimal lymphocytic inflammation, no interface activity and patchy mild bile ductular proliferation highlighted by cytokeratin 7 and cytokeratin 19 immunostains (Figure 4C,D). Bile ducts were missing in some of the portal tracts, but not sufficiently to fulfill the criteria for bile duct paucity (Figure 4E). The hepatocytes were otherwise unremarkable, except for rare foci of subtle hepatocytic cholestasis (Figure 4F). Duodenal and colonic biopsies showed no significant abnormality (Figure 5A,B). At his most recent clinic visit, the proband was 5 years-old and he had started taking ursodeoxycholic acid and fat-soluble vitamins. Although with time his growth had normalized, his coagulopathy had resolved and he had not developed any new liver or gastrointestinal-related signs or symptoms; he continued to have persistently elevated transaminases, direct bilirubin and GGT levels (Table 1).
Table 1. Clinical and laboratory findings in the proband who harbors a homozygous premature termination mutation (p.Gln189*) in SLC51A.

<table>
<thead>
<tr>
<th>Laboratory Test</th>
<th>Reference</th>
<th>Proband’s Age in Years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2.5</td>
</tr>
<tr>
<td>AST (U/L)</td>
<td>15-40</td>
<td>126 (H)</td>
</tr>
<tr>
<td>ALT (U/L)</td>
<td>4-25</td>
<td>227 (H)</td>
</tr>
<tr>
<td>Bilirubin total/direct (mg/dL)</td>
<td>0.1-1.2/0.3</td>
<td>0.7/n.a.</td>
</tr>
<tr>
<td>GGT (U/L)</td>
<td>5-32</td>
<td>n.a.</td>
</tr>
<tr>
<td>ALP (U/L)</td>
<td>100-320</td>
<td>708 (H)</td>
</tr>
<tr>
<td>INR</td>
<td>0.9-1.1</td>
<td>3.1 (H)</td>
</tr>
<tr>
<td>Vitamin D, 25-OH (ng/mL)</td>
<td>16-65</td>
<td>8.7</td>
</tr>
<tr>
<td>Albumin (mg/dL)</td>
<td>3.5-5.5</td>
<td>n.a.</td>
</tr>
<tr>
<td>Total Cholesterol (mg/dL)</td>
<td>&lt;170</td>
<td>n.a.</td>
</tr>
<tr>
<td>HDL (mg/dL)</td>
<td>&gt;35</td>
<td>n.a.</td>
</tr>
<tr>
<td>LDL (mg/dL)</td>
<td>&lt;110</td>
<td>n.a.</td>
</tr>
<tr>
<td>Triglycerides (mg/dL)</td>
<td>32-116</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

**Bile Acids Measured in DBS samples**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>n.a.</th>
<th>n.a.</th>
<th>n.a.</th>
<th>0.02</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cholic acid</td>
<td>0.003-0.19</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.11</td>
</tr>
<tr>
<td>Dihydroxycholanoates* (µM)</td>
<td>0.09-2.21</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.11</td>
</tr>
<tr>
<td>Glycocholic acid (µM)</td>
<td>0.005-2.52</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td></td>
</tr>
<tr>
<td>Glycodihydroxycholanoates** (µM)</td>
<td>0.22-4.16</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.33</td>
</tr>
<tr>
<td>Taurocholic acid (µM)</td>
<td>0.001-0.612</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.06</td>
</tr>
<tr>
<td>Taurodihydroxycholanoates*** (µM)</td>
<td>0.03-0.43</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.07</td>
</tr>
</tbody>
</table>

*on vitamin D supplement; **prior to initiation of ursodeoxycholic acid supplement
*Chenodeoxycholic acid + deoxycholic acid; **Glycochenodeoxycholic acid + glycodeoxycholic acid; ***Taurochenodeoxycholic acid + taurodeoxycholic acid
n.a., not available; H, high; DBS, dry blood spots
Table 2. Summary of laboratory tests performed as part of the comprehensive evaluation for abnormal liver function and chronic diarrhea prior to pursuit of whole-exome sequencing.

<table>
<thead>
<tr>
<th>Labs</th>
<th>Proband’s Value (Normal range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAV, HBV, HCV, CMV, EBV, HIV, VZV, ParvoB19, Rubella, HSV-1, HSV2</td>
<td>negative</td>
</tr>
<tr>
<td>pANCA</td>
<td>negative</td>
</tr>
<tr>
<td>cANCA</td>
<td>negative</td>
</tr>
<tr>
<td>Anti-LKM1</td>
<td>negative</td>
</tr>
<tr>
<td>Anti-smooth muscle antibodies</td>
<td>negative</td>
</tr>
<tr>
<td>Coomb’s test</td>
<td>negative</td>
</tr>
<tr>
<td>Iron (mg/dL)</td>
<td>85 (50–120)</td>
</tr>
<tr>
<td>Ferritin (ng/mL)</td>
<td>229.8 (7–140)</td>
</tr>
<tr>
<td>TIBC (mg/dL)</td>
<td>260 (250–425)</td>
</tr>
<tr>
<td>Alpha-1 antitrypsin (mg/dL)</td>
<td>162 (100-300)</td>
</tr>
<tr>
<td>IgA (g/L)</td>
<td>0.62 g/dL (0.27-1.95)</td>
</tr>
<tr>
<td>IgE (IU/mL)</td>
<td>205.4 (&lt;60)</td>
</tr>
<tr>
<td>IgM (mg/dL)</td>
<td>89 (24-210)</td>
</tr>
<tr>
<td>Anti-tissue transglutaminase IgA (U/mL)</td>
<td>0.9 (&lt;7.0)</td>
</tr>
<tr>
<td>Anti-tissue transglutaminase IgG (U/mL)</td>
<td>3.0 (&lt;6.0)</td>
</tr>
<tr>
<td>Cholesterol (mg/dL)</td>
<td>Total: 140 (&lt;170)</td>
</tr>
<tr>
<td></td>
<td>HDL: 32</td>
</tr>
<tr>
<td></td>
<td>LDL: 93</td>
</tr>
<tr>
<td>Triglyceride (mg/dL)</td>
<td>77 (32-116)</td>
</tr>
</tbody>
</table>
Figure 4. Liver histology of the proband with undiagnosed cholestatic disease. Liver biopsies sectioned from the proband with (A) haematoxylin and eosin (HE) stain showing distorted lobular architecture and expanded portal areas with mild ductular proliferation. (B) Trichrome stain showing portal and periportal fibrosis along with many thin fibrous septa. Nodularity of the lobular architecture and features of early cirrhosis are obvious at this magnification. (C) Cytokeratin 7 immunostain highlights the portal tracts and shows mild bile ductular proliferation in most of the portal areas. (D) Higher magnification of cytokeratin 19 immunostain of one of the portal tracts showing presence of two bile duct profiles (arrows). (E) Two portal areas are depicted; one with few bile ductular profiles (arrow) and one devoid of any bile ducts (circle) The portal areas show only minimal lymphocytic inflammation and no interface activity. Hepatocytes are otherwise unremarkable. (F) Higher magnification of periportal area showing rare foci of subtle hepatocytic cholestasis. Scale Bars shown in lower right corner of each figure, measuring 100 µm in figure 4 A, B, and C and 50 µm in the remaining panels. As published in part in Gao et al. 2019 (55)
Figure 5. Duodenal and colonic mucosa of the proband. (A) H&E stain of duodenal and (B) colonic mucosa showing no significant histological abnormality. Scale Bars, 50 µm. As published in part in Gao et al. 2019 (55)
Homozygous loss-of-function mutation in SLC51A

The proband’s exome was sequenced to a mean depth of 44 independent reads per targeted base, with 96% of targeted bases having more than eight independent reads, providing high-confidence calling of homozygous and heterozygous variants across the exome (Table 3). Calculated inbreeding coefficient of 0.067 was consistent with first-cousin union (Figure 6A). We thus focused our analysis on rare homozygous variants. Consistent with consanguinity, the proband harbors 23 rare homozygous protein-altering variants (Table 4). Twenty-two of these resulted in missense variants. None of these missense variants were predicted to be damaging by MetaSVM and were therefore unlikely to be the cause of disease in this child. The other homozygous mutation led to a premature termination at codon 186 in SLC51A (NM_152672, c.556C>T, p.Gln186*), which encodes the solute carrier organic soluble transporter alpha (OSTα), a known transporter with an important role in bile acid homeostasis (25). This variant allele is extremely rare, with an overall allele frequency in gnomAD of $4 \times 10^{-6}$ with only 1 instance of this variant among more than 251,400 alleles sequenced from diverse populations, including South Asians (Table 4). Sanger sequencing confirmed the homozygous variant in the proband and showed the heterozygous carrier state of both his parents and an unaffected sister (Figure 6B). Eigenstrat showed ethnic clustering of the proband to South Asian populations (Figure 7). The predicted amino acid change from this variant occurs in the fourth of eight transmembrane domains of the OSTα transporter (Figure 8).
Table 3. Quality metrics of proband’s exome.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean independent reads per targeted base</td>
<td>44</td>
</tr>
<tr>
<td>Read length, bp</td>
<td>100</td>
</tr>
<tr>
<td>Percent of bases mapping to genome</td>
<td>99.95</td>
</tr>
<tr>
<td>Percent of targeted bases with ≥ 8 independent reads</td>
<td>93.8</td>
</tr>
<tr>
<td>Mean error rate, %</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Figure 6. Pedigree and Sanger sequencing of the affected proband and his unaffected family members. (A) Pedigree of the proband’s family with affected and unaffected subjects shown in black and white respectively; consanguineous union is represented by a double line. SLC51A alleles are denoted as WT or Mut. (B) Sanger sequencing chromatogram of the nonsense mutation in the proband and his unaffected parents and siblings are shown. The SLC51A mutation at chr3:195955714 C>T (NM_152672), resulting in the 186 codon change from CAA (Glutamine) to TAA (stop codon) is homozygous in the proband and heterozygous in the unaffected parents and sister. Figure as published in Gao et. al. 2019 (55)
Table 4. Rare homozygous protein-altering variants identified in the proband.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chr:Position</th>
<th>AA Change</th>
<th>MetaSVM score prediction(a)</th>
<th>gnomAD (overall)</th>
<th>gnomAD (highest frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLC5A</td>
<td>3:195955714</td>
<td>Gln186*</td>
<td>truncating</td>
<td>3.98E-06</td>
<td>6.152E-05 (African)</td>
</tr>
<tr>
<td>UAP1</td>
<td>1:162557401</td>
<td>Asn324Ser</td>
<td>-1.019 (T)</td>
<td>6.01E-05</td>
<td>3.01E-04 (East Asian)</td>
</tr>
<tr>
<td>MROH2A</td>
<td>2:234698170</td>
<td>Tyr102Asn</td>
<td>-1.071 (T)</td>
<td>1.19E-05</td>
<td>1.20E-04 (South Asian)</td>
</tr>
<tr>
<td>UTRN</td>
<td>6:144750869</td>
<td>Asn283His</td>
<td>-1.036 (T)</td>
<td>2.16E-4</td>
<td>1.76E-03 (South Asian)</td>
</tr>
<tr>
<td>PCM1</td>
<td>8:17868817</td>
<td>Arg504Ser</td>
<td>-0.969 (T)</td>
<td>3.88E-03</td>
<td>6.53E-03 (Non-Finnish European)</td>
</tr>
<tr>
<td>CYP7A1</td>
<td>8:59404935</td>
<td>Pro398Ala</td>
<td>-0.912 (T)</td>
<td>3.36E-03</td>
<td>6.33E-03 (Finnish European)</td>
</tr>
<tr>
<td>BHLHE22</td>
<td>8:65493627</td>
<td>Gly94Ser</td>
<td>-0.084 (T)</td>
<td>1.69E-03</td>
<td>2.84E-03 (Ashkenazi Jewish)</td>
</tr>
<tr>
<td>BICD2</td>
<td>9:95481848</td>
<td>Ala360Val</td>
<td>-1.04 (T)</td>
<td>4.60E-05</td>
<td>1.68E-04 (South Asian)</td>
</tr>
<tr>
<td>FAM208B</td>
<td>10:5790722</td>
<td>Phe1780Val</td>
<td>-0.926 (T)</td>
<td>6.0E-04</td>
<td>4.77E-03 (South Asian)</td>
</tr>
<tr>
<td>SHOC2</td>
<td>10:112724729</td>
<td>Thr205Ala</td>
<td>-0.919 (T)</td>
<td>5.18E-05</td>
<td>3.92E-04 (South Asian)</td>
</tr>
<tr>
<td>NELL1</td>
<td>11:21594825</td>
<td>Ser704Asn</td>
<td>-0.536 (T)</td>
<td>3.39E-04</td>
<td>2.68E-03 (South Asian)</td>
</tr>
<tr>
<td>NCKAP5L</td>
<td>12:50189841</td>
<td>Pro316Leu</td>
<td>-1.098 (T)</td>
<td>1.42E-05</td>
<td>1.23E-04 (South Asian)</td>
</tr>
<tr>
<td>C14orf93</td>
<td>14:23467689</td>
<td>Arg182Trp</td>
<td>-0.818 (T)</td>
<td>1.32E-04</td>
<td>9.63E-04 (South Asian)</td>
</tr>
<tr>
<td>ACIN1</td>
<td>14:23549416</td>
<td>Ser394Arg</td>
<td>-1.063 (T)</td>
<td>4.70E-04</td>
<td>3.69E-03 (South Asian)</td>
</tr>
<tr>
<td>BCL11B</td>
<td>14:99723953</td>
<td>Asp94Glu</td>
<td>-1.03 (T)</td>
<td>5.49E-04</td>
<td>4.48E-03 (South Asian)</td>
</tr>
<tr>
<td>ISL2</td>
<td>15:76630847</td>
<td>His168Pro</td>
<td>-0.288 (T)</td>
<td>2.11E-04</td>
<td>2.53E-03 (South Asian)</td>
</tr>
<tr>
<td>MESDC2(b)</td>
<td>15:81282077</td>
<td>Arg19Gly</td>
<td>Not available</td>
<td>3.26E-04</td>
<td>2.52E-03 (South Asian)</td>
</tr>
<tr>
<td>ZCCHC2</td>
<td>18:60241937</td>
<td>Val875Leu</td>
<td>-1.115 (T)</td>
<td>8.02E-06</td>
<td>6.5E-05 (South Asian)</td>
</tr>
<tr>
<td>LGALS16</td>
<td>19:40151145</td>
<td>Asn138Lys</td>
<td>-0.983 (T)</td>
<td>1.18E-03</td>
<td>5.53E-03 (South Asian)</td>
</tr>
<tr>
<td>SF11</td>
<td>22:31999780</td>
<td>Arg537His</td>
<td>-1.027 (T)</td>
<td>4.02E-05</td>
<td>1.94E-04 (South Asian)</td>
</tr>
<tr>
<td>DOCK11</td>
<td>X:117788680</td>
<td>Gln160Arg</td>
<td>-1.047 (T)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MAGEC3</td>
<td>X:140969340</td>
<td>Pro223Thr</td>
<td>-1.038 (T)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(a\)MetaSVM scores missense variants on a scale of -2 to 3, with scores <0 predicted to be tolerated (T) and scores > 0 predicted to be damaging (D)

\(b\)MESDC2 variants are consecutive nucleotide changes on the same allele and therefore MetaSVM score cannot be calculated. Polyphen HumVar for this variant was predicted to be benign (49). AA, amino acid; Chr, chromosome
Figure 7. Principle component analysis of the proband with homozygous p.Gln186* mutation in SLC51A (light blue star) shows that the proband cluster with individuals of South Asian ancestry from the HapMap project (light blue circles).
Figure 8. Schematic depiction of the location of the proband’s terminating mutation in OSTα. OSTα is a seven-transmembrane protein, depicted in green. The red triangle depicts the location of the terminating mutation in the proband, in the fourth transmembrane subunit. An early termination mutation at this location would result in a non-functional protein, and likely mRNA decay. The second subunit of the OSTα-β transporter, OSTβ, is a single-transmembrane protein, depicted in grey. Expression of both subunits has been shown to be necessary for the functionality of this bile acid transporter.
SLC51A, but not SLC51B, is absent in the proband’s colon

To confirm absence of SLC51A expression in proband tissues, RNA was extracted from FFPE colonic mucosal tissue of the proband and two controls. The controls colons were normal-appearing, from one adult and one child without known liver or colonic disease. While SLC51A and SLC51B transcripts were detected in the normal-appearing colonic mucosal control, the proband’s colon tissue showed only SLC51B expression with undetectable mRNA expression of SLC51A (Figure 9). This pattern of expression was consistent with the above recessive premature termination mutation in SLC51A leading to mRNA decay.

Figure 9 Gene expression of SLC51A and SLC51B relative to GAPDH in proband and control colon tissues. N. D.- not detected. Figure as published in Gao et. al. 2019.
The proband’s blood spot bile acid concentrations were not elevated

Given that the proband had underlying cholestasis while also lacking a major transporter that returns bile acids to the portal systemic circulation and the liver, we quantified serum bile acid levels. Analysis of the bile acid concentrations in the proband’s blood spot by LC-MS/MS, prior to initiation of ursodeoxycholic acid, showed that all bile acid concentrations were within the normal range in spite of underlying cholestasis (Table 1).
DISCUSSION

Our findings provide the first demonstration of human OSTα deficiency in a child with elevated liver transaminases, elevated GGT cholestasis, liver fibrosis and congenital diarrhea. The evidence implicating SLC51A as the cause of the proband’s phenotype is strong.

First, the proband harbors a homozygous mutation for an extremely rare premature termination mutation in SLC51A, representing the first human recessive loss-of-function variant reported in this gene. Second, there is absence of SLC51A mRNA expression in the proband’s colon tissue, consistent with his genotype. Third, from a clinical standpoint, the proband’s history of chronic diarrhea, evidence of coagulopathy and deficiency of fat-soluble vitamins, failure to thrive, cholestasis and elevated hepatic transaminases are all consistent with a defect in bile acid metabolism. OSTα-β is a bi-directional, diffusion-dependent bile acid transporter. While under normal physiological conditions it is thought to reabsorb recycled bile acids from the portal vein to the liver, it has been suggested that under pathological conditions, OSTα-β transporters might be hepatoprotective by exporting bile acids into the plasma to decrease toxic levels in the liver (29, 56). This is supported by the fact that in cholestatic disorders, such as primary biliary cholangitis and biliary obstruction, hepatic expression of OSTα-β is increased (29). Additionally, OSTα-β inhibition has been proposed as a potential mechanism of drug-induced liver injury (DILI). Several drugs associated with cholestatic DILI were shown to inhibit OSTα-β activity in vitro by
greater than 25% (28). Fourth, human deficiency for OSTβ, which forms a heterodimer with OSTα in the active transporter, has been recently described in two siblings who also presented with malabsorptive diarrhea and features of cholestasis (27, 57). Collectively, these three patients with either OSTα deficiency or OSTβ deficiency, all have elevated transaminases (ALT>AST), elevated GGT, and low-normal circulating bile acids.

Aside from the SLC51A variant, the only other rare variant our proband harbors in a gene associated with the bile acid pathway is CYP7A1, which encodes for cholesterol-7-α hydroxylase. This enzyme is responsible for the initial step in bile-acid metabolism (58). However, the CYP7A1 variant harbored by our proband results in an amino acid substitution predicted to be benign by MetaSVM. Furthermore, the phenotype of cholesterol-7-α hydroxylase deficiency is characterized by hypercholesterolemia with an average total cholesterol level above 300 mg/dL among CYP7A1 homozygotes, with no signs of cholestasis, no abnormalities in liver function tests, and no diarrhea (59). In contrast, our proband has normal cholesterol levels (Table 1) and a phenotype most notable for congenital diarrhea, cholestasis, persistently elevated liver function tests, and liver fibrosis (Table 1).

Mouse models of genetic cholestasis do not consistently recapitulate the human disease phenotype due to differences in tissue gene expression and bile acid metabolism between mice and humans (25, 60-62). This is clearly demonstrated
in the OSTα deficient mouse. SLC51A mRNA is abundantly expressed in human liver whereas minimal to no expression is seen in whole livers of mice and rats (25). Additionally, two independent OSTα deficient mouse lines available do not show any liver pathology (63, 64). Although the OSTα-β requires expression of both subunits for stability and function of the transporter, the relative gene expression of each subunit does not occur in a 1:1 ratio in several tissues (25). Notably, in humans, SLC51A mRNA is expressed more abundantly than SLC51B in the liver (25). Hence, it is interesting that our OSTα deficient proband has a more severe liver phenotype as compared to liver histological findings reported in an OSTβ deficient patient (27). Specifically, liver biopsy of one of the OSTβ deficient patients at age 10 showed that the hepatic lobular architecture was intact, with a mild degree of portal fibrosis without features of bile retention or bile duct pathology (27), whereas the liver biopsy of our OSTα deficient child at age 4 showed both bile duct loss and bridging fibrosis and imminent cirrhosis (Figure 4 A, B). Further studies are needed to investigate whether the unique abundance of hepatic SLC51A transcripts as compared to SLC51B may account for a predominant and/or additional role(s) of OSTα in the liver, possibly explaining the difference in liver disease severity seen among human OSTα and OSTβ deficiency.

It is notable that our proband and the OSTβ deficient siblings have normal or near-normal serum bile acid levels in the presence of cholestasis. Usually cholestasis is associated with elevated serum bile levels. One possible
explanation of this is that there are two opposing effects that lead to normal plasma levels in OST\(\alpha\)-\(\beta\) deficiency. Failure of bile acid absorption from the gut may lead to lower blood bile acid concentrations. Decreased bile acid uptake from the blood by the liver may increase blood bile acid concentrations. The net effect may result in normal plasma levels seen in these patients.

A limitation of this study is that our proband is the first patient reported to have OST\(\alpha\) deficiency, and there were only two patients previously reported to have OST\(\beta\) deficiency. With a total of three patients reported to lack the OST\(\alpha\)-\(\beta\) transporter, it is uncertain whether the observed difference in liver fibrosis is a true difference between OST\(\alpha\) deficiency and OST\(\beta\) deficiency, or whether overall the transporter deficiency has a highly variable liver presentation, reflected by the spectrum of disease between the patients. An additional shortcoming of the study was the absence of antibodies to detect OST\(\alpha\) protein expression in control samples. Immunostaining for OST\(\alpha\) was attempted on control slides of liver and intestine, but unsuccessful. We attempted staining with two antibodies (an in-house antibody and a Santa Cruz N15 sc-100078 antibody) with two different antigen retrievals (10 mm citrate pH 6, 1 mm EDTA, pH8). The in-house antibody successfully immunostained OST\(\alpha\) in frozen liver samples in Ballatori et. al. 2005 (25). However, our proband’s biopsies were formalin-fixed paraffin-embedded (FFPE) colon and liver tissue. Our in-house antibodies have never been shown to immunostain FFPE samples, and to our knowledge the Santa Cruz N15 sc-100078 antibodies have not been shown to stain such
samples either. In light of this limitation, we extracted mRNA from formalin-fixed paraffin-embedded (FFPE) colon samples of the proband and the controls and assessed OSTα and OSTβ expression at the mRNA level. While there is known variability in the quality of RNA isolated from FFPE samples (65), we consistently found the absence of SLC51A expression and presence of SLC51B expression in our proband and the expression of both subunits in our control samples, using taqman gene expression assays (Figure 9).

Collectively, our findings highlight the importance of using genomic analysis to study patients with idiopathic disease. Although the SLC51A gene was discovered in Leucoraja erinacea and subsequently in humans over 15 years ago, it was not until genomic analysis was applied to children with undiagnosed liver disease that the first case of human OSTα deficiency was identified. Our study also shows the utility in studying genetics in highly consanguineous communities, and in South Asian communities that have been historically underrepresented in genomics research. Several papers have addressed the particular promise of conducting research in countries such as Pakistan given the high rate of consanguinity and the range of biodiversity within this population (66). The conclusions of our research have enabled us to give a diagnosis to our proband’s previously idiopathic liver fibrosis and congenital diarrhea. Long-term follow-up of the proband and the two siblings with OSTβ deficiency (27) as well as additional patients with these deficiencies is needed to determine the spectrum of clinical features attributable to loss of function of these transporters.
Improved understanding of the clinical course and features of OSTα deficiency through the discovery of additional patients with OSTα–β deficiency may enable us to offer further prognosis and clinical management for our proband.

In summary, this study provides further evidence on how an unbiased genomic approach is highly valuable in many areas of biomedicine (46, 67, 68), including in the field of Hepatology (12-14, 69). In addition to pinpointing the etiology of diseases that elude standard diagnostic evaluations, whole exome sequencing has the power to uncover genetic contributions to human health and disease.
REFERENCES

33. Rahmani S, ABOU AF, Pourbarghi M, DOULATKHAH H, MIRZA AA. The frequency of consanguineous marriages and their effects on offspring’s in Tabriz City. 2010.


