Published in final edited form as:

Clin Genet. 2010 July ; 78(1): 47–56. doi:10.1111/j.1399-0004.2009.01353.x.

Secondary and tertiary structure modeling reveals effects of novel mutations in polycystic liver disease genes *PRKCSH* and *SEC63*

E Waanders^{a,b}, H Venselaar^c, RHM te Morsche^b, DB de Koning^{b,d}, PS Kamath^e, VE Torres^f, S Somlo^g, and JPH Drenth^b

^aDepartment of Human Genetics, and Hepatology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands ^bDepartment of Gastroenterology and Hepatology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands ^cCenter for Molecular and Biomolecular Informatics, Nijmegen Center for Molecular Life Sciences, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands ^dDepartment of Gastroenterology and Hepatology, Jeroen Bosch Hospital, 's-Hertogenbosch, The Netherlands ^eDivision of Gastroenterology and Hepatology, Mayo Clinic College of Medicine, Rochester, MN, USA ^fDivision of Nephrology and Hypertension, Mayo Clinic College of Medicine, Rochester, MN, USA ^gDepartments of Medicine (Nephrology) and Genetics, Yale School of Medicine, New Haven, CT, USA

Abstract

Polycystic liver disease (PCLD) is characterized by intralobular bile duct cysts in the liver. It is caused by mutations in *PRKCSH*, encoding hepatocystin, and *SEC63*, encoding Sec63p. The main goals of this study were to screen for novel mutations and to analyze mutations for effects on protein structure and function. We screened 464 subjects including 76 probands by direct sequencing or conformation-sensitive capillary electrophoresis. We analyzed the effects of all known and novel mutations using a combination of splice site recognition, evolutionary conservation, secondary and tertiary structure predictions, POLYPHEN, and PMUT and SIFT. We identified a total of 26 novel mutations in *PRKCSH* ($n = 14$) and *SEC63* ($n = 12$), including four splice site mutations, eight insertions/deletions, six non-sense mutations, and eight missense mutations. Out of 48 PCLD mutations, 13 were predicted to affect splicing. Most mutations were located in highly conserved regions and homology modeling for two domains of Sec63p showed severe effects of the residue substitutions. In conclusion, we identified 26 novel mutations associated with PCLD and we provide *in silico* analysis in order to delineate the role of these mutations.

© John Wiley & Sons A/S, 2010

Corresponding author: Joost P.H. Drenth, MD, PhD, Department of Gastroenterology and Hepatology, Radboud University Nijmegen Medical Center, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands. Tel.: +31 24 361 4760; fax +31 24 354 0103; JoostPHDrenth@CS.com.

Additional Supporting information may be found in the online version of this article.

Please note: Wiley-Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Keywords

homology modeling; polycystic liver disease; *PRKCSH*; *SEC63*; structural effects

Polycystic liver disease (PCLD; MIM# 174050) is a rare but potentially disabling disease characterized by overgrowth of fluid-filled cysts limited to the liver. The cysts arise from the intralobular bile ducts and are lined with a single layer of biliary epithelial cells (1, 2). The number and size of the cysts may cause severe hepatomegaly with concomitant symptoms such as severe dyspnoea and wasting (3, 4).

PCLD is a congenital disease and is inherited in an autosomal dominant fashion (5, 6). Two groups independently identified *PRKCSH* (MIM# 177060) in 2003 as the molecular culprit causing PCLD. Drenth and colleagues identified two founder mutations among four families, whereas Li and coworkers discovered six mutations in six families (7, 8). The discovery of *PRKCSH* came as a surprise because the function of the encoded protein, the non-catalytic β -subunit of glucosidase II or hepatocystin, is apparently unrelated to known pathways of cystogenesis. Glucosidase II is located in the endoplasmic reticulum (ER) and is involved in carbohydrate processing and quality control of glycoproteins (2, 9). A second locus was mapped to 6q21 following the identification of a family with autosomal dominant PCLD not linked to the *PRKCSH* gene (10). The linkage interval contained an attractive candidate; *SEC63* (MIM# 608648). Indeed, seven mutations were found in this gene in five families and three singleton cases (10). *SEC63* encodes Sec63p, which, like the protein product of *PRKCSH*, is an ER resident protein (2). Sec63p functions in the protein translocation into and out of the ER (11, 12). In all, mutations of either *PRKCSH* or *SEC63* gene account for approximately 25% of PCLD cases, indicating the existence of at least one more locus for PCLD (7, 8, 10, 13).

Mutation analysis of both the *PRKCSH* and the *SEC63* genes may become increasingly important for establishing a diagnosis of PCLD in affected patients. It can be helpful in the differential diagnosis with autosomal dominant polycystic kidney disease (ADPKD) where patients sometimes also have polycystic livers as a predominant feature. This is important as renal function is preserved in PCLD but not in ADPKD.

In an effort to provide a full overview of all known mutations in *PRKCSH* and *SEC63*, we collated results of mutation analyses from institutions currently involved in genetic research in PCLD. We expected to shed light on the molecular pathology of PCLD and to aid the development of an effective molecular testing strategy for diagnostic testing. Therefore, we established two objectives: (i) to find novel mutations by screening the coding regions of *PRKCSH* and *SEC63* in patients presented for PCLD DNA diagnostics and (ii) to estimate the possible effect of new and described mutations on the structure and function of the proteins using bioinformatics methods.

Materials and methods

Patients

For this study we used data from two diagnostic centers: the Radboud University Nijmegen Medical Center and the Mayo Clinic-Yale University collaborative group. In these centers, molecular diagnostics is carried out on patient samples from all over Europe and North America. Blood samples presented on suspicion of PCLD between 2004 and 2008 were screened for mutations in *PRKCSH* and *SEC63*. In Nijmegen, we analyzed 214 singleton cases and nine unrelated probands from a total of 26 samples. In the United States, mutation analysis was carried out in 67 unrelated affected probands from a total of 214 samples. A proband was defined as a patient with at least one affected family member. PCLD was diagnosed in probands or isolated individuals when >20 liver cysts were present. In case of a positive family history, patients aged 40 years or younger with any liver cysts or patients above the age of 40 with four or more liver cysts were diagnosed with PCLD. None of the individuals have been previously reported and all patients gave an informed consent.

Genomic DNA extracted from peripheral blood (Gentra Systems, Minneapolis, MN) was used for mutation screening through direct sequencing or conformation-sensitive capillary electrophoresis (CSCE).

Direct sequencing

Direct sequencing was performed as described earlier (13). Briefly, exons and adjoining intronic segments including splice donor and acceptor sites of *PRKCSH* and *SEC63* were amplified using standard polymerase chain reaction (PCR) (primer sequences available upon request). The PCR products were verified on gel electrophoresis and subsequently purified using QIAEXII Gel Extraction Kit (Qiagen, Hilden, Germany). Finally, the samples were sequenced using the BigDye terminator kit and an ABI3730 capillary sequencer (Applied Biosystems, Foster City, CA).

Conformation-sensitive capillary electrophoresis

The heteroduplex-based mutation detection consisted of two standard PCR reactions with primers complementary to flanking intronic sequences (available upon request). The first forward primer contained a 5' M13 primer sequence and the second forward primer was a M13 forward primer labeled with Fam, Ned, Vic, or Rox. After PCR, the products were pooled (one of each color) and diluted in MQ containing a GeneScan 500 LIZ size standard (Applied Biosystems, Foster City, CA). After purification, the samples were run on an ABI 3730 with a 48-capillary array (Applied Biosystems, Foster City, CA). Finally, the acquired data were analyzed using the GeneMapper software (Applied Biosystems, Foster City, CA). Deviant samples were sequenced to determine the exact mutation.

Mutation nomenclature

To determine primer and mutation positions, we used *PRKCSH* genomic DNA (GenBank Accession No. NC_000019.8), cDNA (GenBank Accession No. NM_002743.2), and protein sequences (GenBank Accession No. NP_002734.2) as well as *SEC63* genomic DNA (GenBank Accession No. NC_000006.10), cDNA (GenBank Accession No. NM_007214.3),

and protein sequences (GenBank Accession No. NP_009145.1). Mutation nucleotide numbering was based on the cDNA sequences with an A of the ATG translation initiation codon designated as nucleotide +1. All mutations are described according to mutation nomenclature standards (www.hgvs.org/mutnomen) (14).

***In silico* analysis**

All mutations were tested for influence on (cryptic) splice sites using the web-based splice site prediction by neural network (NNSPLICE09; http://www.fruitfly.org/seq_tools/splice.html) (15).

Subsequently, the effect of amino acid substitutions resulting from missense mutations was analyzed using different approaches. We examined the conservation of the proteins using multiple protein sequence alignments constructed with the ENSEMBL genome browser (www.ensembl.org), Clustal-W, and Jalview (www.jalview.org) (16). Hepatocystin was aligned with 19 orthologues and Sec63p with 17 orthologs (Table S1, supporting information online). Orthologs were selected on availability, validity (no predicted proteins), and range of species (including fish, birds, rodents, mammals, and primates). The *PRKCSH* gene contains a large helix structure and we analyzed mutations in this region using a helical wheel predictor (<http://cti.itc.virginia.edu/~cmg/Demo/wheel/wheelApp.html>). We were able to perform homology modeling for two domains of Sec63p: the luminal DnaJ domain and SEC63-2 domain (amino acids 626–719). The model of the human J-domain was based on the J-domain from mouse DnaJ subfamily C (PDB-file 2CTW, 45% identity over 75 amino acids). The model for the SEC63_2 domain was based on PDB-file 2Q0Z (38% identity over 93 amino acids). Homology modeling and subsequent analysis were performed using Yasara's homology modeling experiment (17). This method fully automatically performs the model building process including alignment generation, loop building, rotamer selection, optimization, and validation of the model. The obtained overall Z-scores for DnaJ domain and the SEC63-2 domain fell in the range 'good' and "satisfactory," respectively. More information, including Yasara's modeling reports, can be found online: www.cmbi.ru.nl/~hvensela/PRKCSH/. Subsequent visualization and analysis was performed using the Yasara & WHAT IF Twinset.

To assess the effect of the mutations in the remaining part of the proteins, we used a secondary structure prediction using the PsiPredserver (PSIPRED v2.6, <http://bioinf.cs.ucl.ac.uk/psipred/>) (18, 19). Finally, we analyzed all missense mutations using three different pathogenic-or-not predictors. These predictors integrate several analyses to come to a prediction on pathogenicity of the mutation (see for a review on possible analyses; (20)). We used PolyPhen (<http://genetics.bwh.harvard.edu/pph/>), pMut (<http://mmb2.pcb.ub.es:8080/PMut/>), and sift (<http://blocks.fhcrc.org/sift/SIFT.html>) (21–23).

Results

PCLD mutation screening

We conducted an extensive screen for mutations in both *PRKCSH* and *SEC63* of up to 464 samples including 76 unrelated probands. In addition to the previously reported 13 *PRKCSH*

and nine *SEC63* mutations, we found 14 hitherto unreported variants in *PRKCSH* and 12 in *SEC63* (Tables 1 and 2, Fig. 1). All mutations were found in a heterozygous state. The missense variants were rare variants, i.e. they were found in less than 1% of the tested population (>1000 alleles), and are therefore regarded as pathogenic mutations (26). One patient was compound heterozygous for both *PRKCSH* and *SEC63* mutations, but showed no exceptional phenotype. The mutations are published on www.livercyst.org.

Mutation spectrum

Analysis of the mutation locations revealed that the mutations are distributed over the entire reading frame (Fig. 1b,e). Truncating mutations accounted for 78% of the *PRKCSH* mutations and 62% of the mutations found in *SEC63* (Fig. 1c,f). Truncating mutations were defined as deletions, insertions, splice site mutations, and non-sense mutations that cause a frameshift and/or early termination of protein translation.

Splice site recognition

Next, we analyzed all mutations for effect on (cryptic) splice sites using NNSLIPCE09. All splice site mutations resulted in abolishment of the existing splice site (Table S2, supporting information online). In addition, the non-sense mutation c.487C>T in *PRKCSH* also abolished the neighboring acceptor splice site, even though the mutation is situated 20 base pairs downstream in the exon. The deletion of one base pair in exon 4 of *SEC63* (c.422delT) resulted in the origination of a new splice donor site with a probability score of 0.98. However, it is unclear that whether this site will be preferably used above the original splice donor site, which scores 0.99. None of the other mutations found in this study influenced the strength of the neighboring splice site or resulted in the formation of a new splice site.

Sequence conservation

Furthermore *in silico* analyses focused on the missense mutations. We analyzed hepatocystin and Sec63p from, respectively, 19 and 17 species. We found that both proteins were highly conserved. The alignments showed that the amino acids of 12 out of 14 missense or in frame mutations were conserved between all or almost all species (Table 3).

Homology modeling

We were able to perform homology modeling for two domains of Sec63p: the luminal DnaJ domain and the SEC63-2 domain. Both p.I120T and p.D168H mutations are located in the luminal DnaJ domain, which mainly consists of α -helices (Fig. 2a). The side chain of isoleucine (p.I120T) is located in the hydrophobic space between two helices. Isoleucine is hydrophobic and is important for the stability of the helical domain. Threonine is smaller and less hydrophobic, which might result in loss of the interactions in the core between the helices and consequently loss of stability. The second mutation in the DnaJ domain is an aspartic acid that changes into the bigger neutral histidine (p.D168H) (13). The bigger side chain of histidine does not fit in the space between the helices and is likely to disturb the local conformation.

The second domain in Sec63p that we modeled (amino acids 626–719) contained the mutation p.W651G (Fig. 2b). The large tryptophan at position 651 is located in the core

between two β -sheets. A mutation into a glycine (with no side chain) results in a big hole in the middle of the domain, which alters the structure profoundly and may influence the protein function.

Secondary structure modeling

Secondary structure modeling predicted the *PRKCSH* c.1141G>A (p.E381K) mutation to be located in an α -helix. Using a helical wheel predictor, we showed that this helix is amphipathic, containing hydrophobic and hydrophilic sides (Fig. 3). The hydrophobic side can interact with other domains of the protein, leaving the hydrophilic side free for complex formation with other proteins or maybe another part of the protein. A mutation from the negatively charged glutamic acid into a positively charged lysine might disturb these interactions. Unfortunately, no structure or modeling template was found for *PRKCSH* which complicated tertiary structure modeling.

The *PRKCSH* p.T261S mutation is located in the loop between two helices in the EF hand domain. The small but effective difference between threonine and serine and the location between two helices is reminiscent of a calcium-binding domain. The p.M175V is located in a helix. Helices often interact with each other by hydrophobic side chains and therefore this residue might be important for the packing between two helices. Both methionine and valine are hydrophobic but a change from a large side chain to a smaller β -branched one might lead to less hydrophobic interactions and destabilization of the domain.

In the cytoplasmic domain of Sec63p just downstream of the DnaJ domain, the arginine of p.R217C is predicted to be located in a helix. As arginine is positively charged and can make hydrogen bonds, it is likely that a mutation to a hydrophobic and smaller cysteine disrupts the interaction with other residues. Both p.A373Gdel3aa and p.Q375P in *SEC63* are located in the same region of the protein and are predicted to be in a helix. A substitution of glutamine with a proline (p.Q375P) will have severe impact on the structure of this helix as the torsion of proline will disturb the helix. The fact that an in frame deletion of three amino acids at this location results in PCLD indicates an important functional role for this region of the protein. The p.E568del mutation was previously described by Davila and coworkers (10). This mutation is located in the middle of an α -helix. Deletion of the glutamic acid will result in a shift of all following residues of approximately 100° in the turn of the helix. As a result, all properties of the remaining residues in the helix will move to a different position, which can disturb the structure and function severely.

The secondary structure predictions of hepatocystin and Sec63p are available as Figs S1 and S2, supporting information online, respectively. Detailed information about the *in silico* analyses can be found at www.cmbi.ru.nl/~hvensela/PRKCSH.

Pathogenic-or-not predictors

Finally, we analyzed all missense mutations using three different internet based pathogenic-or-not prediction programs (Table 3). PolyPhen uses structural information as well as multiple sequence alignments for its prediction. This program classified 9 out of 12 missense mutations as possibly or probably damaging. In contrast, pMut uses sequence-

based information only for its prediction and identified 5 out of 12 pathological mutations. SIFT uses phylogenetic information and found 7 of 12 pathogenic. Three mutations in *PRKCSH* were classified as benign in all three prediction programs, even though we here provide explanations for their pathogenicity based on sequence conservation and 2D modeling.

Discussion

Here, we describe 26 hitherto unreported mutations together with an extensive structural analysis of all 48 mutations described in *PRKCSH* and *SEC63*. We found that all splice site mutations, one nonsense mutation, and one single-base pair deletion influenced splice site recognition. Furthermore, most missense mutations are located in conserved regions and have a severe impact on secondary or tertiary protein structure. We were able to give a possible explanation for the mutation pathogenicity for all mutations now identified in association with PCLD.

Excluding family members of the probands, we found 62 mutations in a total of 299 samples. This results in a mutation frequency of 20.7% (15% in *PRKCSH* and 5.7% in *SEC63*) indicating that at least one other locus is involved in PCLD pathogenesis. However, in PCLD research, an estimation of mutation frequencies has some limitations. First, the diagnostic criteria of PCLD are ambiguous. In general, the lower limit of 20 liver cysts is used for PCLD diagnosis in isolated cases, but we found mutations in two patients who carried <20 liver cysts. In addition, we previously showed that patients with as few as eight liver cysts may carry mutations in *PRKCSH* and *SEC63* (13). Second, because the disease is often asymptomatic or non-penetrant, family history is frequently negative. This might explain the large number of singleton cases in our cohort. In our study, we found the *PRKCSH* mutations c.292+1G>C and c.1341-2A>G in, respectively, eight and 11 unrelated singleton cases. However, Drenth and colleagues have previously shown that both mutations occur in the Netherlands due to a founder effect (7). Therefore, when calculating mutation frequencies, haplotype analyses should always be used to exclude family ties and founder effects. Third, we see numerous patients with a large polycystic liver in combination with a few or multiple kidney cysts, but without renal failure. These patients comply to the Ravine criteria for ADPKD and are excluded from PCLD mutation frequency studies (27). However, an overlapping phenotype between ADPKD and PLCD is not inconceivable. This hypothesis is supported by a recent study that showed an interaction between PCLD protein hepatocystin (*PRKCSH*) and ADPKD protein polycystin-2 (*TRPP2*) (28).

All missense mutations were found in less than 1% of the studied population (>1000 alleles). Cotton and coworkers stated that a mutation that occurs in less than 1% of the population is by definition rare and probably pathogenic (26). In contrast, neutral variants are likely to escape negative evolutionary selection and are thus more prevalent. In concordance, even though some of the new mutations described here showed a low level of conservation, lacked a severe difference in amino acid characteristics, or are not predicted pathogenic in our *in silico* analyses, they are expected to be disease-causing mutations. As it is unclear how mutations in hepatocystin and Sec63p lead to cystogenesis in PCLD, the identification

of important protein domains by mutational analysis might reveal the mechanism of molecular functions, protein–protein interactions or protein–glycan interactions.

We identified one patient who was transheterozygous for the common *PRKCSH* c.1341-2A>G mutation and the novel *SEC63* c.359T>C (p.I120T) mutation. To our knowledge, this is the only patient described with mutations in both PCLD genes. Unfortunately, family information was not available and we could not investigate whether the *SEC63* c.359T>C mutation alone can result in PCLD. However, our *in silico* analyses very convincingly showed the pathogenic effects of the *SEC63* mutation.

The in frame deletion of glutamic acid (p.E568del) in *SEC63* was identified in four unrelated female patients. Interestingly, even though the analyses showed that the mutation is very deleterious, we found a marked variation in disease phenotype. Two of the patients displayed a severe phenotype, with more than 20 liver cysts and massive hepatomegaly at the age of 39 and 43 (liver volumes of 6.3 l). In contrast, the other two patients presented with a few small cysts on radiological studies at the age of 63 and 82 (two cysts of 2.5 cm and one cyst of 3 cm, respectively). This suggests that the penetrance of the disease is indeed less than 100%. On the other hand, these results can also indicate that other genetic or environmental factors play a role in disease initiation and/or progression.

Finally, it was recently described that not only promoter methylation regulates transcription of a gene, but intragenic methylation also has influence (29, 30). Therefore, the disruption or creation of CpG dinucleotides by mutations may alter the transcription regulation of a gene. In this study, we found six missense mutations located at CpG dinucleotides, three of which were C>T transitions. Cytosine to thymidine mutations are most common at CpG dinucleotides, as a methylated cytosine is receptive to spontaneous deamination resulting in a thymidine (31).

In conclusion, 48 mutations have now been identified to be involved in PCLD pathogenesis, and their effect on protein structure was evident. Nevertheless, 79% of the patients screened did not show a mutation in either *PRKCSH* or *SEC63* indicating involvement of at least one more locus.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We sincerely thank Prof G Vriend for his expert advice and critical reading of the manuscript. Joost PH Drenth is supported by a VIDI fellowship from the Netherlands Organization for Scientific Research (NWO).

References

1. Desmet VJ. Congenital diseases of intrahepatic bile ducts: variations on the theme “ductal plate malformation”. *Hepatology*. 1992; 16:1069–1083. [PubMed: 1398487]
2. Waanders E, Croes HJ, Maass CN, et al. Cysts of *PRKCSH* mutated polycystic liver disease patients lack hepatocystin but express *Sec63p*. *Histochem Cell Biol*. 2008; 129:301–310. [PubMed: 18224332]

3. Tahvanainen P, Tahvanainen E, Reijonen H, et al. Polycystic liver disease is genetically heterogeneous: clinical and linkage studies in eight Finnish families. *J Hepatol.* 2003; 38:39–43. [PubMed: 12480558]
4. Qian Q, Li A, King BF, et al. Clinical profile of autosomal dominant polycystic liver disease. *Hepatology.* 2003; 37:164–171. [PubMed: 12500201]
5. Pirson Y, Lannoy N, Peters D, et al. Isolated polycystic liver disease as a distinct genetic disease, unlinked to polycystic kidney disease 1 and polycystic kidney disease 2. *Hepatology.* 1996; 23:249–252. [PubMed: 8591848]
6. Drenth JP, Martina JA, van de Kerkhof R, Bonifacino JS, Jansen JB. Polycystic liver disease is a disorder of cotranslational protein processing. *Trends Mol Med.* 2005; 11:37–42. [PubMed: 15649821]
7. Drenth JP, Te Morsche RH, Smink R, et al. Germline mutations in PRKCSH are associated with autosomal dominant polycystic liver disease. *Nat Genet.* 2003; 33:345–347. [PubMed: 12577059]
8. Li A, Davila S, Furu L, et al. Mutations in PRKCSH cause isolated autosomal dominant polycystic liver disease. *Am J Hum Genet.* 2003; 72:691–703. [PubMed: 12529853]
9. Trombetta ES, Fleming KG, Helenius A. Quaternary and domain structure of glycoprotein processing glucosidase II. *Biochemistry.* 2001; 40:10717–10722. [PubMed: 11524018]
10. Davila S, Furu L, Gharavi AG, et al. Mutations in SEC63 cause autosomal dominant polycystic liver disease. *Nat Genet.* 2004; 36:575–577. [PubMed: 15133510]
11. Meyer HA, Grau H, Kraft R, et al. Mammalian Sec61 is associated with Sec62 and Sec63. *J Biol Chem.* 2000; 275:14550–14557. [PubMed: 10799540]
12. Ponting CP. Proteins of the endoplasmic-reticulum-associated degradation pathway: domain detection and function prediction. *Biochem J.* 2000; 351(Pt 2):527–535. [PubMed: 11023840]
13. Waanders E, Te Morsche RH, de Man RA, et al. Extensive mutational analysis of PRKCSH and SEC63 broadens the spectrum of polycystic liver disease. *Hum Mutat.* 2006; 27:830. [PubMed: 16835903]
14. den Dunnen JT, Paalman MH. Standardizing mutation nomenclature: why bother? *Hum Mutat.* 2003; 22:181–182. [PubMed: 12938082]
15. Reese MG, Eeckman FH, Kulp D, et al. Improved splice site detection in Genie. *J Comput Biol.* 1997; 4:311–323. [PubMed: 9278062]
16. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007; 23:2947–2948. [PubMed: 17846036]
17. Krieger E, Koraimann G, Vriend G. Increasing the precision of comparative models with YASARA NOVA – a selfparameterizing force field. *Proteins.* 2002; 47:393–402. [PubMed: 11948792]
18. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol.* 1999; 292:195–202. [PubMed: 10493868]
19. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics.* 2000; 16:404–405. [PubMed: 10869041]
20. Thusberg J, Vihinen M. Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. *Hum Mutat.* 2009; 30:703–714. [PubMed: 19267389]
21. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* 2002; 30:3894–3900. [PubMed: 12202775]
22. Ferrer-Costa C, Orozco M, de la Cruz X. Sequence-based prediction of pathological mutations. *Proteins.* 2004; 57:811–819. [PubMed: 15390262]
23. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003; 31:3812–3814. [PubMed: 12824425]
24. Drenth JP, Tahvanainen E, Te Morsche RH, et al. Abnormal hepatocystin caused by truncating PRKCSH mutations leads to autosomal dominant polycystic liver disease. *Hepatology.* 2004; 39:924–931. [PubMed: 15057895]
25. Peces R, Drenth JP, Te Morsche RH, et al. Autosomal dominant polycystic liver disease in a family without polycystic kidney disease associated with a novel missense protein kinase C substrate 80K-H mutation. *World J Gastroenterol.* 2005; 11:7690–7693. [PubMed: 16437702]

26. Cotton RG, Scriver CR. Proof of “disease causing” mutation. *Hum Mutat.* 1998; 12:1–3. [PubMed: 9633813]
27. Ravine D, Gibson RN, Walker RG, et al. Evaluation of ultrasonographic diagnostic criteria for autosomal dominant polycystic kidney disease 1. *Lancet.* 1994; 343:824–827. [PubMed: 7908078]
28. Gao H, Wang Y, Wegierski T, et al. PRKCSH/80K-H, the protein mutated in polycystic liver disease, protects polycystin-2/TRPP2 against HERP-mediated degradation. *Hum Mol Genet.* 2009 Oct 3. Epub ahead of print. [PubMed: 19801576]
29. Lorincz MC, Dickerson DR, Schmitt M, et al. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat Struct Mol Biol.* 2004; 11:1068–1075. [PubMed: 15467727]
30. Nguyen C, Liang G, Nguyen TT, et al. Susceptibility of non-promoter CpG islands to de novo methylation in normal and neoplastic cells. *J Natl Cancer Inst.* 2001; 93:1465–1472. [PubMed: 11584062]
31. Shen JC, Rideout WM III, Jones PA. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.* 1994; 22:972–976. [PubMed: 8152929]

Fig. 1.

Mutation analyses of *PRKCSH* -hepatocystin and *SEC63* -Sec63p. (a) The location of the different mutations depicted on the schematic structure of hepatocystin. The structure consists of an N-terminal signaling sequence (yellow), low-density lipoprotein receptor domain (LDLa, orange), two putative calcium-binding EF-Hand domains (blue), a glutamic acid repeat (pink), a variable protein domain (green), a mannose-6-phosphate receptor domain (purple), and the C-terminal HDEL ER-retention signal sequence (yellow). Mutations found in this study are highlighted with a red asterisk. (b) Mutation location in *PRKCSH*. Notice the large number of mutations in exon 6. (c) Types of mutations found in hepatocystin. Missense mutations are found in a small fraction (22%, light blue). Truncating mutations (78%, dark blue) are predominant in hepatocystin and consist of deletions (15%, green), insertions (15%, yellow), non-sense mutations (22%, purple), and splice site mutations (26%, red). (d) The location of the different mutations depicted on the schematic structure of Sec63p. The integral ER-membrane protein contains three trans-membrane spanning domains (green tubes), a luminal DnaJ-like domain between trans-membrane

segments 2 and 3 (*red*) and a large cytoplasmic domain with a negatively charged C-terminus (–). Mutations found in this study are highlighted with a red asterisk. (e) Mutation location in *SEC63*. The mutations are equally divided over the exons. (f) Mutation types found in Sec63p. Missense mutations and in frame deletions are found in almost half of the cases (38%, *light blue* and *orange*). Truncating mutations (62%, *dark blue*) consist of deletions (14%, *green*), insertions (5%, *yellow*), non-sense mutations (24%, *purple*), and splice site mutations (16%, *red*).

Fig. 2.

Homology modeling of Sec63p. The models are shown in gray with the side chain of the residues exposed. The mutation sites are visible with wild-type residues in green and mutations in red. (a) Model for the DnaJ domain. The p.D168H mutation is depicted on the left. Notice the large mutant side chain that does not fit in the space between the helices. The p.I120T mutation is depicted on the right. The smaller side chain of the mutation may result in the loss of stability between two adjacent helices. (b) Model for the C-terminal 626–719 amino acids of Sec63p. The p.W651G mutation location is visible in green for the wild-type situation. The mutant glycine has no side chain and therefore leaves a hole in the structure.

Fig. 3. Helical wheel prediction of hepatocystin. The p.E381K mutation changing glutamic acid into a lysine is located in a helix. The helix has a hydrophobic and hydrophilic side. A mutation from the negatively charged glutamic acid (position 10) into a positively charged lysine may disturb interactions of the helix with other domains. The helical plot was constructed using amino acids 369–395 of the hepatocystin protein sequence.

Table 1

PRKCSH mutations identified in PCLD patients

Position	Nucleotide change	Amino acid change	No. of alleles ^a	Origin	Reference
<i>Splice site</i>					
IVS 2	c.76-79+4dup8	p.N27SfsX117	0		(13)
IVS 4	c.292+1G>C	p.D98AfsX100	8	Dutch	(7)
IVS 8	c.684-4delGCAG	p.V229X	2	USA	This study
IVS 9	c.762+2T>C	p.L292PfsX330	0		(8)
IVS 15	c.1341-2A>G	p.T448X	11	Dutch	(7)
IVS 15	c.1341-1G>A	p.T448X	1	USA	This study
IVS 16	c.1440+1 delGT	p.T448CfsX457	1	Finnish	(8)
<i>Insertion/deletion</i>					
Exon 4	c.215_216insA	p.N72KfsX81	0		(8)
Exon 6	c.353_354insA	p.K119EfsX122	1	USA	This study
Exon 6	c.368delA	p.E123GfsX130	1	USA	This study
Exon 6	c.374_375delAG	p.E125VfsX145	4	Dutch	(24)
Exon 6	c.430_432delCTTins7bp	p.L144NfsX147	3	Dutch	This study
Exon 8	c.668delA	p.D223VfsX231	1	USA	This study
Exon 13	c.1168_1169insC	p.I391HfsX401	0		(8)
Exon 15	c.1336delC	p.G447AfsX463	2	USA	This study
<i>Non-sense</i>					
Exon 6	c.466C>T	p.Q156X	1	USA	This study
Exon 7	c.487C>T	p.Q163X	2	UK	This study
Exon 7	c.593G>A	p.W198X	1	USA	This study
Exon 14	c.1240C>T	p.Q414X	0		(8)
Exon 15	c.1269C>G	p.Y423X	0		(8)
Exon 16	c.1395T>G	p.Y465X	2	Dutch	This study
<i>Missense</i>					
Exon 6	c.416G>A	p.R139H	1	Dutch	(13)
Exon 6	c.464A>G	p.K155R	0		(13)
Exon 7	c.523A>G	p.M175V	1	USA	This study

Position	Nucleotide change	Amino acid change	No. of alleles ^a	Origin	Reference
Exon 10	c.781A>T	p.T261S	1	French Canadian	This study
Exon 10	c.841C>T	p.R281W	0		(25)
Exon 13	c.1141G>A	p.E381K	1	USA	This study
Total number of mutations found in this study			45		

^aNumber of alleles found in this study excluding family members. Mutations identified in this study are depicted in bold face.

Table 2

SEC63 mutations identified in PCLD patients

Position	Nucleotide change	Amino acid change	No. of alleles ^a	Origin	Reference
<i>Splice site</i>					
IVS 1	c.125-2A>G	p.E42GfsX156	0		(10)
IVS 8	c.733+1G>A	p.L246KfsX247	0		(10)
IVS 8	c.733+1G>T	p.L246KfsX247	1	French	This study
IVS 11	c.1053_1054+4delAGgtga	p.R353EfsX355	1	Turkish	This study
<i>Insertion/deletion</i>					
Exon 4	c.422delT	p.M141SfsX143	1	Dutch	This study
Exon 4	c.441_442insA	p.A148SfsX155	0		(10)
Exon 12	c.1118_1126del9bp	p.A373G del 3aa	1	Dutch	This study
Exon 17	c.1702_1704delGAA	p.E568del	4	Dutch	(10)
Exon 17	c.1813_1817delCAAAA	p.N606RfsX607	1	Dutch	This study
Exon 19	c.2006_2007delAAT	p.H669RfsX689	0		(10)
<i>Non-sense</i>					
Exon 2	c.173G>A	p.W58X	0		(10)
Exon 3	c.292C>T	p.R98X	1	Dutch	This study
Exon 8	c.715C>T	p.R239X	1	Dutch	This study
Exon 10	c.891T>A	p.Y297X	0		(10)
Exon 16	c.1577C>A	p.S526X	0		(13)
<i>Missense</i>					
Exon 4	c.359T>C^b	p.I120T^b	1^b	Dutch	This study
Exon 5	c.502G>C	p.D168H	0		(13)
Exon 8	c.649C>T	p.R217C	1	Dutch	This study
Exon 9	c.801A>C	p.R267S	2	USA	This study
Exon 12	c.1124A>C	p.Q375P	1	Dutch	This study
Exon 19	c.1951T>G	p.W651G	1	Dutch	This study
Total number of mutations found in this study			17		

^a Number of alleles found in this study excluding family members.

^b Mutation found transheterozygously with *PRKCSH* c.1341-2A>G. Mutations identified in this study are depicted in bold face.

Table 3
Analysis of missense mutations in *PRKCSH* and *SEC63* of PCLD patients

Gene	Nucleotide change	Amino acid change	Conserved in 19 species ^a	Structure prediction	PolyPhen	pMut	SIFT	Protein domain
<i>PRKCSH</i>	c.416G>A	p.R139H	Yes, fish different		Possibly damaging	Neutral	Tolerated	At CpG
	c.464A>G	p.K155R	Highly		Benign	Neutral	Tolerated	
	c.523A>G	p.M175V	No	2D	Benign	Neutral	Tolerated	
	c.781A>T	p.T261S	No	2D	Benign	Neutral	Tolerated	EF-Hand 2
	c.841C>T	p.R281W	Yes, frog same class; fish different		Possibly damaging	Pathological	Affected	At CpG C>T; EF-Hand 2
<i>SEC63</i>	c.1141G>A	p.E381K	Yes, mouse different	3D	Possibly damaging	Neutral	Affected	At CpG C>T; mannose-6-phosphate receptor
	c.359T>C^b	p.I120T^b	Highly	3D	Probably damaging	Neutral	Affected	Luminal DnaJ
	c.502G>C	p.D168H	Highly	3D	Probably damaging	Neutral	Affected	Luminal DnaJ
	c.649C>T	p.R217C	Yes, fish same class	2D	Probably damaging	Pathological	Affected	At CpG C>T
	c.801A>C	p.R267S	Yes, horse same class		Probably damaging	Pathological	Affected	
In frame deletion	c.1124A>C	p.Q375P	Highly	2D	Probably damaging	Pathological	Affected	Creates CpG
	c.1951T>G	p.W651G	Highly	3D	Probably damaging	Pathological	Affected	
	c.1118_1126del9bp	p.A373G del 3aa	Yes	2D	Probably damaging	Pathological	Affected	Deletes CpG
	c.1702_1704delGAA	p.E568del	Yes	2D				

^aSec63p was aligned with orthologs from 17 species.

^bMutation found transheterozygously with *PRKCSH* c.1341-2A>G. 3D prediction based on homology modeling. 2D prediction based on secondary structure modeling. Mutations identified in this study are depicted in bold face.