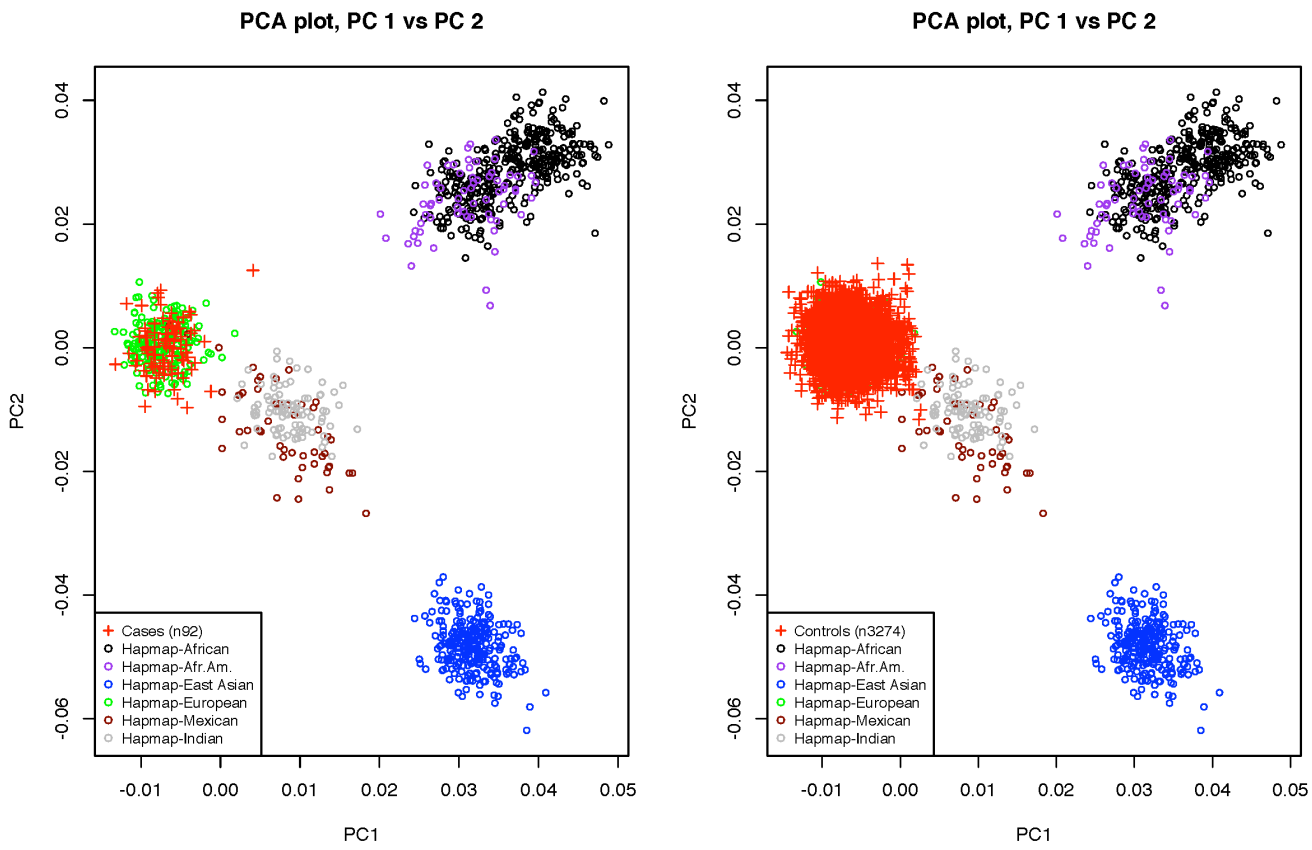
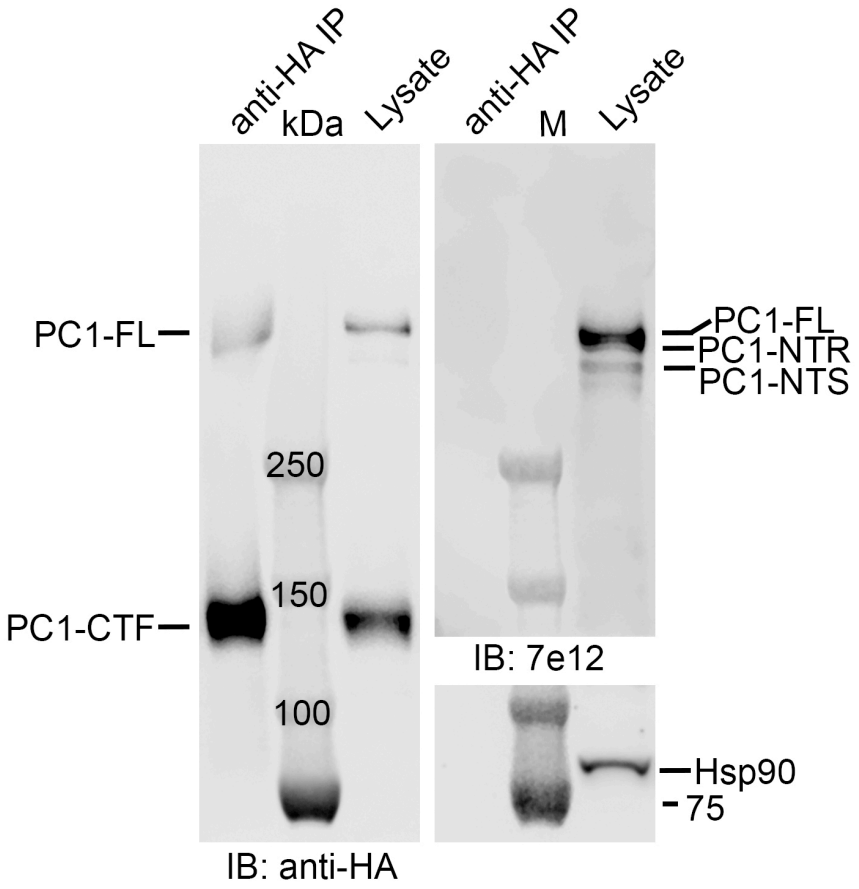
Isolated polycystic liver disease genes define effectors of polycystin-1 function

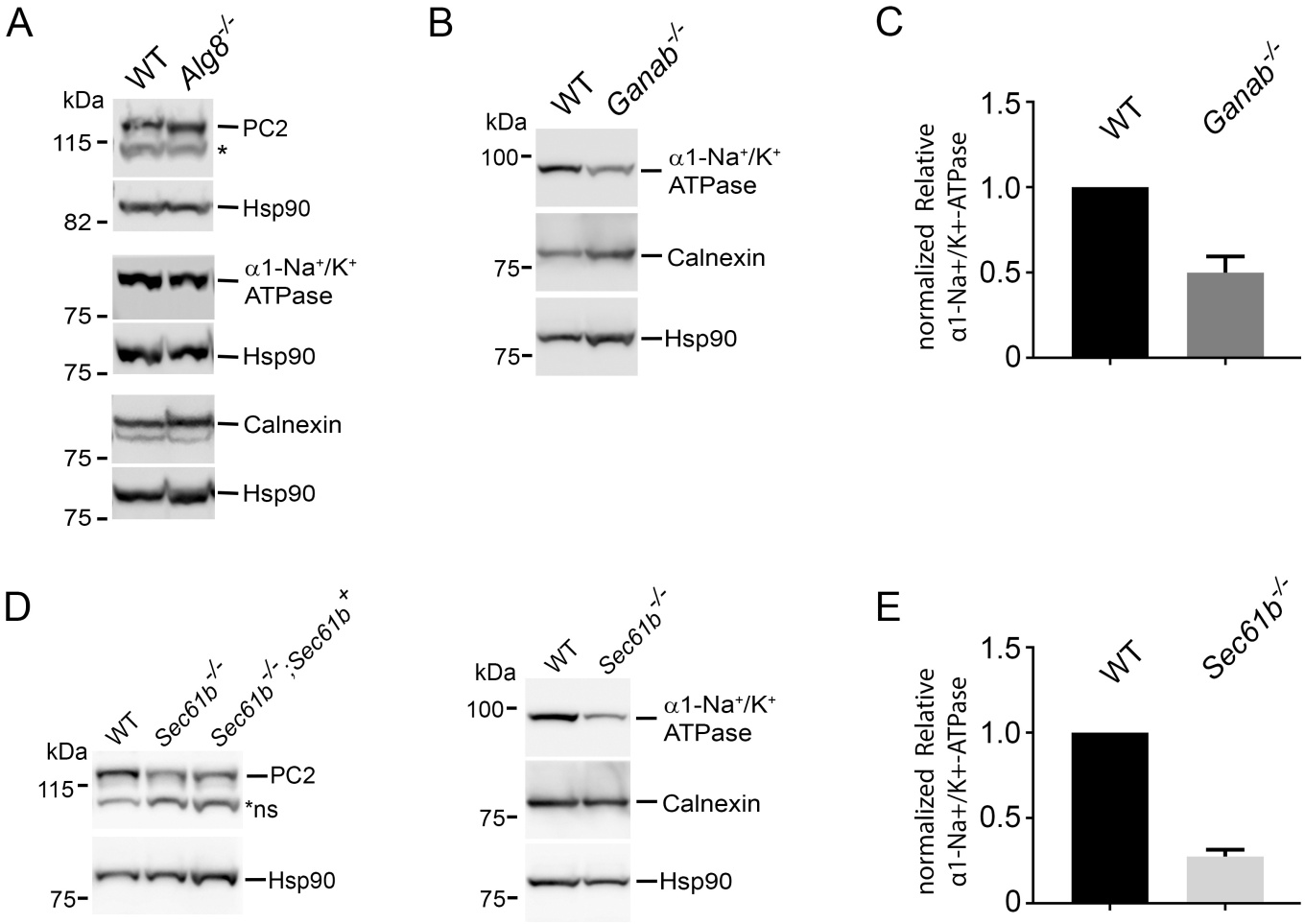
Whitney Besse[1], Ke Dong[1], Jungmin Choi[2]*, Sohan Punia[1]*, Sorin V. Fedeles[1]*, Murim Choi[2], Anna-Rachel Gallagher[1], Emily B. Huang[1], Ashima Gulati[1], James Knight[2], Shrikant Mane[2], Esa Tahvanainen[3], Pia Tahvanainen[3], Simone Sanna-Cherchi[4], Richard P. Lifton[2], Terry Watnick[5], York P. Pei[6], Vicente E. Torres[7], Stefan Somlo[1,2,‡]



Supplemental Figure 1: Principle component analysis (PCA) performed on exome data demonstrating clustering of **(A)** European discovery cohort cases (n = 92) and **(B)** European controls (n = 3274) over established European samples.



Supplemental Figure 2: Native anti-LRR PC1 antibody does not recognize the dual epitope tagged PC1 expressed by the *Pkd1$^{F/H}$-BAC* transgene. Immunoblot with anti-HA (left panel) or anti-LRR (7e12, right top panel) of cells expressing the *Pkd1$^{F/H}$-BAC* transgene that were either immunoprecipitated with anti-HA (anti-HA IP) or used as untreated lysate (Lysate). The same blot was re-probed with both antibodies. Anti-HA shows similar intensity of the PC1-FL protein from the *PKD1$^{F/H}$-BAC* in both HA-IP and Lysate. Reblot of membrane with the anti-LRR N-terminal antibody (7e12) shows absence of signal for PC1-FL in the HA-IP lane. This suggests that the anti-LRR native protein antibody (7e12) is only able to detect the native PC1 protein, not the N-terminal FLAG epitope tagged PC1 expressed by the *Pkd1$^{F/H}$-BAC* transgene. The N-terminal FLAG tag may interfere with 7e12 binding to its epitope.



Supplemental Figure 3: Steady state protein levels of other membrane proteins. (**A**) Immunoblots of PC2, α1-Na^+/K^+-ATPase and calnexin show no difference in *Alg8*$^{-/-}$ cells. *, non-specific band. (**B**) Immunoblots of α1-Na^+/K^+-ATPase and calnexin show reduced levels of α1-Na^+/K^+-ATPase in *Ganab*$^{-/-}$ cells. (**C**) Quantitation of α1-Na^+/K^+-ATPase steady-state protein level normalized to Hsp90 on immunoblots of lysates from *Ganab*$^{-/-}$ and wild type (WT) cells (n=3, P=0.0344). (**D**) Immunoblots of PC2, α1-Na^+/K^+-ATPase and calnexin in *Sec61b*$^{-/-}$ cells. (**E**) Quantitation of α1-Na^+/K^+-ATPase steady-state protein level normalized to Hsp90 on immunoblots of lysates from *Sec61b*$^{-/-}$ and wild type (WT) cells (n=4, P=3x10^{-4}). Paired sample T-test was used for all statistical calculations in this figure, and error bars represent SEM.

Supplemental Table 1: Pathogenic mutations (*n* = 53) in *PRKCSH* and *SEC63* in the PCLD cohort

Proband	Nucleotide change[A]	Amino Acid Change	MAF in ExAC	Reference
PRKCSH				
T-99	c.52A>T	p.K17X	Novel	
T-25[B]	c.215_216insA	p.N72fs	Novel	(1)
T-72	c.352_353insA	p.E118fs	Novel	(2)
T-51, T-79, T-86, T-103, T-105	c.374_375delAG	p.E125fs	8.2×10^{-6}	
T-54	c.466C>T	p.Q155X	Novel	(2)
YU236	c.593G>A	p.W197X	Novel	(2)
T-66	c.667_668delA	p.D223fs	Novel	(2)
T-41[B]	c.762+2T>C	Splice donor	2.0×10^{-5}	(1)
O-1[B], T-53, YU203	c.1168_1169insC	p.N390fs	Novel	(1)
T-2[B]	c.1240C>T	p.Q414X	Novel	(1)
T-56	c.1341-1G>A	Splice acceptor	Novel	(2)
T-88, YU304	c.1267_1268insA	p.R423fs	Novel	
T-7[B], T-98, YU287, YU231	c.1269C>G	p.Y423X	8.2×10^{-6}	(1)
T-47	c.1335delC	p.T444fs	Novel	(2)
YU315	c.1440+1G>A	Splice donor	Novel	
T-1[B], FINN50, FINN63, FINN69, FINN71, FINN72	c.1440+1_1440+2delGT	Splice donor	5.0×10^{-5}	(1)
SEC63				
YU278	c.19C>T	p.Q7X	Novel	
YU193, YU195, YU218	c.173G>A	p.W58X	Novel	(3)[C]
T-27[B]	c.225-2A>G	Splice site	Novel	(3)
T-97	c.292C>T	p.R98X	8.2×10^{-5}	
FINN4[B]	c.441_442insA	p.A148fs	Novel	(3)
T-107, YU251	c.452+1G>A	Splice site	Novel	
T-108	c.699C>G	p.Y233X	Novel	
T-92	c.715C>T	p.R239X	8.9×10^{-6}	
T-52	c.883_884insA	p.C295fs	Novel	
A-6[B]	c.891T>A	p.Y297X	Novel	(3)
YU321	c.1074_1076delACCinsCTAGAG	p.T359X	Novel	
T-96	c.1103_1104delA	p.K368fs	Novel	
T-94	c.1222delAA	p.T408fs	Novel	
T-6	c.1249C>T	p.E417X	Novel	
T-113	c.1481insG	p.E494fs	Novel	
T-78, YU324	c.1577_1578insA	p.K529fs	Novel	
T-114	c.1648A>T	p.K550X	Novel	
YU154	c.1801C>T	p.Q601X	Novel	

[A]Variants were found by either Sanger sequencing or whole exome sequencing; all were confirmed with Sanger sequencing.
[B]Variant always segregated with affected family members (n): T-25 (2), T-41 (1), O-1 (1), T-2 (16), T-7 (3), T-1 (13), T-27 (4), FINN4 (8), A-6 (7).
[C]Variant published previously but not identified as these families.

Supplemental Table 2: Non-synonymous substitution and non-frame shifting deletion mutations in *PRKCSH*, *SEC63*

Proband	Nucleotide change	Amino Acid Change	MAF in ExAC	CADD	Protein Domain[B]
PRKCSH					
T-100	c.154_156delAAC	p.N52del	Novel	-	*PRKCSH*-like domain
SEC63					
YU323	c.185G>A	p.R62Q	8.25×10^{-6}	18.0	Cytoplasmic loop
TOR6468, W-YU217	c.1702_1704del GAA	p.E568del	3.28×10^{-3} [(A)]	-	SEC63 domain

[A]MAF in the Exome Aggregation Consortium (ExAC) database of 4×10^{-3} in Europeans and 1×10^{-2} in Finnish Europeans.
[B]Domain predictions were done using Simple Modular Architecture Research Tool (SMART) Heidelberg, Germany (URL://http://smart.embl-heidelberg.de/) [accessed 6/2016](4). Location prediction using Protter, Zurich, Germany (URL: http://wlab.ethz.ch/protter/start/) [accessed 6/2016].

Supplemental Table 3: Genetic background by cohort determined in exome sequenced probands

	PRKCSH (n=13)	***SEC63*** (n=16)	**Discovery Cohort** (n=102)
European	100% (n=13)	75% (n=12)	90% (n=92)
African American	-	12.5% (n=2)	3% (n=3)
Mexican/Native American	-	12.5% (n=2)	3% (n=3)
Asian	-	-	4% (n=4)

Supplemental Table 4: Exome sequencing quality statistics for discovery cohort cases and European controls

Category	Cases (n=102)		Controls (n=3274)	
	Mean	Median	Mean	Median
Read length (bp)	74.71	74.00	73.82	74.00
Number of reads per sample (M)	73.64	70.70	86.65	76.75
Median independent reads at each targeted base (X)	72.11	68.30	85.21	75.30
Mean independent reads at each targeted base (X)	61.21	58.00	72.17	63.00
Percent of targeted bases with ≥8 independent reads	95.08%	95.50%	95.40%	95.70%

Supplemental Table 5: Sequencing coverages of *PKHD1, ALG8, GANAB, SEC61B* in European discovery cohort cases (n=92) and European controls (n=3274)

Category	*PKHD1*				*ALG8*			
	Cases		Controls		Cases		Controls	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Median independent reads at each targeted base (X)	78.41	76.90	96.29	85.30	110.16	104.50	131.24	119.95
Mean independent reads at each targeted base (X)	70.27	68.50	86.45	76.00	102.07	97.50	120.96	111.00
Percent of targeted bases with ≥8 independent reads	99.02%	99.90%	99.85%	100.00%	96.08%	96.10%	96.03%	96.10%

Category	*GANAB*				*SEC61B*			
	Cases		Controls		Cases		Controls	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Median independent reads at each targeted base (X)	68.39	60.00	68.39	60.00	68.39	60.00	68.39	60.00
Mean independent reads at each targeted base (X)	67.30	57.00	67.30	57.00	67.30	57.00	67.30	57.00
Percent of targeted bases with ≥8 independent reads	99.87%	100.00%	99.87%	100.00%	99.87%	100.00%	99.87%	100.00%

Supplemental Table 6: CRISPR/Cas9 mutant cell lines[A]

Cell Line	Parental Cell Line	gRNA sequence	Cell genotype
***Alg8*$^{-/-}$**	**P**	GCGTCCGGGTCTGCAACCGC (exon 1 forward strand)	compound heterozygous frameshifts: c.23_24insT (p.T8fs), c.23delC (p.T8fs)
***Alg8*$^{-/-}$; *Xbp1*$^{-/-}$**	***Alg8*$^{-/-}$**	GAAAGCCCGGATGAGCGAGC (exon 2 forward strand)	homozygous c.268_271delGAGC (p.E90fs)
***Ganab*$^{-/-}$**	**P + stable expression of Cas9**	#1: GGGGCTGAGGAAATCGGGTG (non-canonical, exon12) #2: GGGGCTGGTGCTGGCCAGGT (non-canonical, exon13)	homozygous 281 bp genomic deletion with frameshift (63 bp of exon 12, all of intron 12, and all 127 bp of exon 13 into intron 13) c.1390_1580del (p.D464fs) deleted exonic sequence: GACCCCACCCGATTTCCTCAGCCCCTCAATATG CTTGAGCACTTGGCTTCCAAGAGGCGGAAGCTGGTGGCCATTGT GGACCCCCACATCAAGGTAGACTCTGGCTACCGAGTTCACGAAG AATTGCGAAACCATGGGCTGTATGTTAAAACTCGGGATGGCTCT GATTACGAGGGCTGGTGCTGGCCAG
***Sec61b*$^{-/-}$**	**P + stable expression of Cas9**	#1: GATGTGGCCTAAACTAACGC (5' to predicted promoter) #2: CGGGATCCACTGTTCGGCAG (forward strand, exon 2)	homozygous 732bp genomic deletion, with random 7 bp insertion (Deletion includes promotor and all of exons 1 and 2) c.1_101del (p.M1_R34del) resultant sequence: TCGGTACAGGTTTTCGGCCACTTACCTCT GGGGGCTTGCGGGGGACG
***Pkhd1*$^{-/-}$ #1**	**P**	TCAAACTGGAAACTCGAGTA (exon 3 reverse strand)	compound heterozygous frameshifts: c.63_64insT (p.S22fs) c.58_68delCCTTACTCGAG (p.P20fs)
***Pkhd1*$^{-/-}$ #2**	**P**	TCAAACTGGAAACTCGAGTA (exon 3 reverse strand)	homozygous frameshift: c.63_64insT (p.S22fs)

[A]All mutations introduced into mouse kidney epithelial cell lines (**P**) conditionally immortalized using the ImmortoMouse transgene [CBA;B10-Tg(H2Kb-tsA58)6Kio/Crl] and containing the *Pkd1*$^{F/H}$*-BAC* transgene for expression of dual epitope tagged PC1 (5, 6).

Supplemental Table 7: Novel PCLD gene heterozygous nonsynonymous or non-frameshifting insertions/deletions of undetermined pathogenicity in PCLD patients in the discovery cohort[A]

Proband	Nucleotide change[B]	Amino Acid Change	MAF in ExAC	PhyloP[E]	CADD score	MetaSVM[F]	ARPKD Database[G]
GANAB							
YU313[C]	c.2614C>T	p.H872Y	$3.28x10^{-3}$	7.5	14.2	T	
R2	c.2419C>A	p.H807N	$8.24x10^{-5}$	2.3	14.9	D	
PKHD1							
YU165	c.12110T>C	p.L4037P	$1.15x10^{-4}$	0.1	15.4	T	-
T-77	c.11338C>T	p.P3780S	$1.57x10^{-3}$	3.7	17.4	D	indeterminate
T-93	c.9788T>C	p.V3263A	$1.73x10^{-3}$	3.4	17.3	T	probably benign (7)
W-YU350[D]	c.9629C>G	p.S3210C	$2.14x10^{-4}$	4.6	17.2	D	-
TOR6467	c.9071G>A	p.C3024Y	Novel	5.2	14.4	T	-
TOR2400[D]	c.7942G>A	p.G2648S	$4.19x10^{-3}$	2.6	17.0	D	indeterminate
T-4[C]	c.7307C>T	p.T2436I	$3.87x10^{-4}$	3.7	20.0	D	-

[A]MAF $<5x10^{-3}$ and CADD score >14.
[B]All variants were found by whole exome sequencing and confirmed with Sanger sequencing.
[C]African American
[D]Asian
[E]PhyloP score representing nucleotide conservation from comparison of human sequence with 99 other vertebrates. (8).
[F]MetaSVM prediction (9): T=Tolerated, D=Deleterious
[G]Mutation Database Autosomal Recessive Polycystic Kidney Disease (ARPKD/PKHD1). Department of Human Genetics, RWTH Aachen University, Aachen, Germany. (URL: http://www.humgen.rwth-aachen.de/) [accessed 6/2016]

Supplemental Table 8: Nonsynonymous substitution mutations in *LRP5* of indeterminate significance

Proband	Nucleotide change	Amino Acid Change	MAF in ExAC	CADD	Protein Domain[A]
LRP5					
T-59	c.1912A>G	p.K638E	$1.65x10^{-5}$	15.38	EGF-like domain
TOR6467	c.2051T>C	p.V684A	Novel	16.71	LDLR YWTD domain
T-58	c.2773C>T	p.R925C	$4.27x10^{-5}$	17.29	EGF-like domain
TOR2400	c.4622C>T	p.T1541M	$1.67x10^{-5}$	16.63	n.r.

[A]Domain predictions were done using Simple Modular Architecture Research Tool (SMART) Heidelberg, Germany (URL://http://smart.embl-heidelberg.de/) [accessed 11/2016](4). n.r = no domain recognized.

References

1. Li A, Davila S, Furu L, Qian Q, Tian X, Kamath PS, King BF, Torres VE, and Somlo S. Mutations in PRKCSH cause isolated autosomal dominant polycystic liver disease. *American journal of human genetics.* 2003;72(3):691-703.

2. Waanders E, Venselaar H, te Morsche RH, de Koning DB, Kamath PS, Torres VE, Somlo S, and Drenth JP. Secondary and tertiary structure modeling reveals effects of novel mutations in polycystic liver disease genes PRKCSH and SEC63. *Clinical genetics.* 2010;78(1):47-56.

3. Davila S, Furu L, Gharavi AG, Tian X, Onoe T, Qian Q, Li A, Cai Y, Kamath PS, King BF, et al. Mutations in SEC63 cause autosomal dominant polycystic liver disease. *Nature genetics.* 2004;36(6):575-7.

4. Letunic I, Doerks T, and Bork P. SMART: recent updates, new developments and status in 2015. *Nucleic acids research.* 2015;43(Database issue):D257-60.

5. Fedeles SV, Tian X, Gallagher AR, Mitobe M, Nishio S, Lee SH, Cai Y, Geng L, Crews CM, and Somlo S. A genetic interaction network of five genes for human polycystic kidney and liver diseases defines polycystin-1 as the central determinant of cyst formation. *Nature genetics.* 2011;43(7):639-47.

6. Fedeles SV, So JS, Shrikhande A, Lee SH, Gallagher AR, Barkauskas CE, Somlo S, and Lee AH. Sec63 and Xbp1 regulate IRE1alpha activity and polycystic disease severity. *The Journal of clinical investigation.* 2015;125(5):1955-67.

7. Gunay-Aygun M, Tuchman M, Font-Montgomery E, Lukose L, Edwards H, Garcia A, Ausavarat S, Ziegler SG, Piwnica-Worms K, Bryant J, et al. PKHD1 sequence variations in 78 children and adults with autosomal recessive polycystic kidney disease and congenital hepatic fibrosis. *Molecular genetics and metabolism.* 2010;99(2):160-73.

8. Pollard KS, Hubisz MJ, Rosenbloom KR, and Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome research.* 2010;20(1):110-21.

9. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, and Liu X. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Human molecular genetics.* 2015;24(8):2125-37.