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 and Stefan Somlo1,2

1Department of Internal Medicine, and 2Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA. 3Department of Medical Genetics, University of Helsinki, Helsinki, Finland. 4Department of Medicine, Columbia University, New York, New York, USA. 5Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA. 6Division of Nephrology, University Health Network, Toronto, Ontario, Canada. 7Division of Nephrology and Hypertension, Mayo Clinic, Rochester, Minnesota, USA.

Dominantly inherited isolated polycystic liver disease (PCLD) consists of liver cysts that are radiologically and pathologically identical to those seen in autosomal dominant polycystic kidney disease, but without clinically relevant kidney cysts. The causative genes are known for fewer than 40% of PCLD index cases. Here, we have used whole exome sequencing in a discovery cohort of 102 unrelated patients who were excluded for mutations in the 2 most common PCLD genes, PRKCSH and SEC63, to identify heterozygous loss-of-function mutations in 3 additional genes, ALG8, CANAB, and SEC61B. Similarly to PRKCSH and SEC63, these genes encode proteins that are integral to the protein biogenesis pathway in the endoplasmic reticulum. We inactivated these candidate genes in cell line models to show that loss of function of each results in defective maturation and trafficking of polycystin-1, the central determinant of cyst pathogenesis. Despite acting in a common pathway, each PCLD gene product demonstrated distinct effects on polycystin-1 biogenesis. We also found enrichment on a genome-wide basis of heterozygous mutations in the autosomal recessive polycystic kidney disease gene PKHD1, indicating that adult PKHD1 carriers can present with clinical PCLD. These findings define genetic and biochemical modulators of polycystin-1 function and provide a more complete definition of the spectrum of dominant human polycystic diseases.

Introduction

Isolated polycystic liver disease (PCLD; also called ADPLD) presents in adult life with radiologically and pathologically identical liver cysts to those seen in autosomal dominant polycystic kidney disease (ADPKD), but without the burden of kidney cysts required for a diagnosis of ADPKD (1–4). Liver cysts in both PCLD and ADPKD are lined with biliary epithelium. Cysts may arise from ductal plate malformations known as von Meyenburg complexes, which result from abnormalities in late embryonic biliary genesis (2, 5, 6), although animal model data show that liver cysts in ADPKD and PCLD can occur in adult life without developmentally determined biliary tract malformations (7, 8). While the inheritance pattern and adult onset of isolated liver cysts in PCLD parallel those of ADPKD, clinical detection may be reduced because liver cysts are often asymptomatic or nonpenetrant (1, 2). Symptomatic PCLD results if the liver size becomes large enough to cause mass effects including abdominal distension, pain, early satiety, and, rarely, biliary or portal obstruction. Symptoms tend to be more severe in women (2, 9). Infrequently, symptoms can be severe enough to require cyst fenestration, partial hepatectomy, or total hepatectomy with liver transplant (10).

Familial linkage-based studies and positional cloning identified autosomal dominant inheritance of loss-of-function mutations in PRKCSH and SEC63 as causing PCLD (1, 11–13). The respective protein products, glucosidase IIß (GIIß) and SEC63, play essential roles in co- and posttranslational modification of membrane and secreted proteins in the endoplasmic reticulum (ER). This was an unexpected localization for polycystic disease gene products since most fibrocystic diseases of the kidney and liver, including ADPKD, are associated with proteins that function in primary cilia (14). Primary cilia are minute solitary membrane-enclosed microtubule-based projections that serve as organelles integrating cellular sensory signals on the apical surface of many cell types, including bile duct and kidney tubule epithelial cells (15–17). Mutations in another noncilial protein, LDL receptor–related protein 5 (LRP5), a coreceptor in canonical Wnt signaling, have also been implicated as causing PCLD (18).

ADPKD, the most common monogenic kidney disease, results in age- and genotype-dependent progression to end-stage renal disease. It is caused by mutations in PKD1 and PKD2, which encode polycystin-1 (PC1) and polycystin-2 (PC2), respectively. PC1 and PC2 are large complex integral membrane proteins that are functionally expressed on primary cilia. PC1 and PC2 are among the many client proteins of GIIß and SEC63 in the ER protein biogenetic pathway. Although inherited as dominant traits,
This hypothesis was validated in vivo by showing that orthologous cystin protein dosage that results in cyst formation in PCLD (13). It was hypothesized that loss of function in either GJB or SEC63 results in impaired biogenesis and maturation of PC1 and PC2, and it is this reduction in polycystin protein dosage that results in cyst formation in PCLD (13). This hypothesis was validated in vivo by showing that orthologous gene animal models with conditional inactivation of Sec63 and Prkcs develop both kidney and liver cysts in a PC1 dosage–dependent manner (7). These studies established that the PCLD genes are modifiers of PC1 function, that inadequate effective functional PC1 underlies cyst formation in PCLD, and that PC1 dosage is the rate-limiting determinant of occurrence and severity for both PCLD and ADPKD (7, 23).

Mutations in PRKCSH and SEC63 explain approximately 35% of human PCLD. We analyzed 102 probands in whom we did not find mutations in either of these genes by whole exome sequencing to identify rare heterozygous loss-of-function mutations as candidate genes for PCLD. No predominant gene emerged, suggesting substantial genetic heterogeneity in the remaining PCLD patient group. In a minority of patients, we did find enrichment of heterozygous carrier mutations in PKHD1, the causative gene for autosomal recessive polycystic kidney disease (ARPKD), and heterozygous inactivating mutations in ALG8, GANAB, and SEC6GB, which all encode ER proteins involved in the same pathway as GJB and SEC63. We used functional bioassays of PC1 biogenesis and trafficking to show that inactivation of these ER-associated PCLD candidate genes caused defective maturation and trafficking of PC1 whereas inactivation of PKHD1 had no such effect. Our findings extend the genetic determinants of PCLD to 4 more genes, and with that we have causative mutations in an additional 15% of our cohort. The findings define essential steps in the biogenesis of mature PC1 protein and extend the hypothesis that PCLD occurs due to reduced effective PC1 function. By extension, the finding that PKHD1 carriers can present with PCLD provides evidence in humans for the hypothesis that the function of the PKHD1 gene product fibrocystin may intersect with polycystin signaling in vivo (7, 23, 24).

Results

Description of cohorts. We studied a total cohort of 159 unrelated individuals with clinically defined PCLD in which each proband had at least 10 liver cysts and did not fulfill the diagnostic criteria for ADPKD (4). Twenty-eight of the 159 individuals had previously defined mutations in either of 2 known genes, PRKCSH and SEC63 (Supplemental Table 1; supplemental material available online with this article; https://doi.org/10.1172/JCI90129DS1). The remaining 131 individuals underwent whole exome sequencing, and an additional 25 individuals with loss-of-function mutations in PRKCSH or SEC63 were identified (Supplemental Table 1). Four other probands had non-frameshifting deletions or non-synonymous substitution variants of uncertain pathogenicity in SEC63 or PRKCSH (Supplemental Table 2). After excluding the 53 probands with PCLD due to PRKCSH or SEC63 and the 4 individuals with the indeterminate variants, we arrived at a discovery cohort of 102 unrelated individuals for finding additional PCLD genes. We compared characteristics of this discovery cohort and the PRKCSH/SEC63 group. Affected family members were available for 14 of 57 (25%) of the PRKCSH/SEC63 probands and for 15 of 102 (15%) of the unknown cases. In keeping with the increased clinical severity of PCLD in female patients, which likely leads to enhanced ascertainment, 31 of 41 (76%) of the PRKCSH/SEC63 group and 74 of 102 (73%) of the discovery cohort were female. Principal component analysis (PCA) comparing all cases that had whole exome sequencing (n = 131) with individuals of defined ancestry from HapMap345 was used to define the ethnic background of PCLD probands in both the PRKCSH/SEC63 and discovery cohorts (Supplemental Table 3). Overall, the discovery and PRKCSH/SEC63 PCLD groups were largely comparable.

Analysis of the discovery cohort. We searched for novel PCLD genes in the discovery cohort by whole exome sequence analysis. We confirmed the unrelated status, defined as identity by descent < 0.125, of each proband in this group using pairwise comparison of exome data. We prospectively specified as candidates for PCLD all genes with loss-of-function variants (i.e., nonsense, splice site, or frameshift) that had a minor allele frequency (MAF) less than 1 × 10^–5 in the Exome Aggregation Consortium (ExAC) database. The MAF threshold was chosen conservatively because of uncertainty about the actual prevalence of radiographically defined PCLD. Autopsy studies have suggested prevalence ranging from 1:400 to 1:2,000, with the largest study showing 1:1,300 (25–27). We found 1,266 sequence variants meeting these criteria in the 102 probands. None of these variants occurred in homozygous or compound heterozygous states, suggesting that PCLD in this cohort is largely a heterozygous trait.

We conducted 2 analyses to evaluate the significance of the per-gene burden of these rare loss-of-function variants. First, a binomial test was used to compare the expected and observed number of variants in each gene. The expected number of loss-of-function variants meeting our criteria was determined by taking the fractional length of a gene relative to the entire exome (i.e., the coding length of the transcript divided by the number of bases in the total exome) and multiplying this by 1,266 (the total number of variants). This represents the expected occurrence of sporadic mutations of this type in each gene without considering the influences of “hot spots,” pressures of selection, or precedents of ethnic background. A quantile-quantile (Q-Q) plot comparing the distribution of observed and expected probabilities for all genes containing loss-of-function variants in our discovery cohort shows genome-wide significant divergence from the expected for only 2 genes, PKHD1 and ALG8, which encodes a glucosyltransferase (Figure 1A and Table 1). Genome-wide significance threshold was taken as P less than 2.4 × 10^–5 after accounting for examination of 18,715 genes. The expected number of loss-of-function variants in PKHD1 among the 1,266 total variants was 0.48, but the observed number was 9 (P = 2.31 × 10^–4; 19-fold enrichment; Table 1). The expected number of loss-of-function variants in ALG8 was 0.06, and the observed number was 5 (P = 7.18 × 10^–4; 81-fold enrichment; Table 1).
We next performed a case-versus-control analysis of gene burden for the loss-of-function variants in the European subset of our discovery cohort (n = 92) versus an available cohort of 3,274 European normal controls sequenced and analyzed by the Yale Center for Mendelian Genomics (28–30). PCA analysis confirmed clustering of the 92 cases and 3,274 controls over European standards (Supplemental Figure 1). The exome coverage metrics met accepted standards, and there was no significant difference in these metrics between cases and controls (Supplemental Tables 4 and 5). A Q-Q plot comparing the observed distribution of Fisher’s exact statistics for the frequency of loss-of-function variants in cases versus controls shows genome-wide significance for heterozygous loss-of-function variants only in PKHD1 (P = 1.55 × 10–9) (Figure 1B and Table 1). The gene with the next lowest P value, ALG8 (P = 2.65 × 10–6), just missed the threshold for genome-wide significance (Figure 1B and Table 1). Only 4 of 5 ALG8 loss-of-function variants occurred in European cases, yielding 4 of 184 European case alleles compared with 1 of 6,548 European control alleles (Tables 1 and 2). The remaining ALG8 loss-of-function allele occurred in an African American individual. Analysis of a similarly obtained but smaller African American control group (392 individuals, 784 alleles) found no loss-of-function ALG8 alleles. In aggregate, both analyses support PKHD1 and ALG8 as candidate genes for PCLD.

We noted that the next-highest-scoring gene in the case-versus-control analysis was GANAB (P = 2.01 × 10–5) (Figure 1 and Tables 1 and 2). GANAB encodes the α subunit of GII (GII α), which is the catalytic subunit that works with the previously defined PCLD gene product GII β (encoded by PRKCSH), to form the func-

**Table 1. Enrichment of rare heterozygous loss-of-function variants in PKHD1, ALG8, GANAB, and SEC61B in PCLD**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein name</th>
<th>Binomial analysis*</th>
<th>European case vs. control†</th>
</tr>
</thead>
<tbody>
<tr>
<td>PKHD1</td>
<td>Fibrocystin/Polyductin</td>
<td>Observed (n)</td>
<td>Expected</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG8</td>
<td>α-1,3-Glucosyltransferase</td>
<td>5</td>
<td>0.06</td>
</tr>
<tr>
<td>GANAB</td>
<td>Glucosidase II α</td>
<td>3</td>
<td>0.11</td>
</tr>
<tr>
<td>SEC61B</td>
<td>SEC61β</td>
<td>1</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Only loss-of-function variants (frameshift, nonsense, splice site) with MAF less than 1 × 10–3 are included, except as noted. All variants were verified by direct Sanger sequencing of amplicons. No individual among either cases or controls had homozygous loss-of-function variants in these genes. †Total discovery cohort, n = 102. P value is determined by the binomial test. ‡European discovery cohort cases, n = 92 (n = 184 alleles). European controls, n = 3,274 (n = 6,548 alleles). P value is calculated using Fisher’s exact test comparing allele counts. §Includes the p.T36M mutation, which is a founder mutation in Europeans and is the most common pathogenic variant for ARPKD (17% of mutant alleles) (41). MAF for this variant is 8.55 × 10–4 in non-Finnish Europeans in ExAC. It is present in 1 case in the discovery cohort and 4 members of the control population. Since this variant does not strictly meet our definition for loss-of-function variants, we provide P values for PKHD1 both excluding and including the p.T36M variant. ¶Only 4 of 5 cases in our discovery cohort with loss-of-function variants in ALG8 are European. The fifth case is African American. ‖Includes a non-synonymous substitution variant, p.M1V, which is predicted to cause loss of translation since there is no alternative in-frame initiation codon. This variant is not present in 3,274 European controls. ExAC has 1 allele with a different sequence change that also results in loss of the start codon (p.M1L, 1 of 121,342 alleles).
tional GII holoenzyme in the ER lumen. Two of the three GANAB variants in our cohort are novel, and the third is exceedingly rare, with MAF less than $1 \times 10^{-5}$ (Table 2). In the analysis based on the binomial distribution, GANAB had $P = 2.25 \times 10^{-4}$ with a predicted burden of 0.11 and an enrichment of 26-fold in the 3 observed cases (Table 1). GANAB is a highly loss-intolerant gene, with a probability of loss-of-function intolerance score (pLI) of 1 in ExAC, whereas PKHD1 and ALG8 are not loss intolerant, consistent with genes associated with a recessive human disease (31). To evaluate GANAB further, we performed a follow-up analysis using the binomial test including only genes with very rare loss-of-function variants (MAF $< 5 \times 10^{-5}$) and pLI greater than 0.9 in our PCLD cohort. The observed distribution for all genes conforms closely to the expected with the exception of GANAB, which diverged significantly. The expected number of loss-of-function alleles for GANAB was 0.02, while the observed number was 3 ($P = 1.46 \times 10^{-6}; 143$-fold enrichment).

Finally, we noted a novel loss-of-function mutation in SEC61B, encoding a component of the SEC61 ER membrane protein-conducting channel that associates with SEC63 (32). There are no loss-of-function mutations for SEC61B in ExAC. We found a second patient with a novel, likely loss-of-function missense variant in SEC61B that mutated the initiation codon, leaving no apparent in-frame alternative start site for translation (Table 2). While it did not reach statistical significance on a genome-wide level, SEC61B showed 87-fold change in the binomial analysis, and, if both variants are considered in the case-versus-control analysis, SEC61B had $P = 7.43 \times 10^{-4}$ (Tables 1 and 2).

Further independent validation of these analyses was done using the European non-Finnish and Finnish combined subset of the population in the ExAC database. Gene-by-gene comparison showed enrichment of mutations for each of our candidate genes in PCLD cases: PKHD1, $P = 2.34 \times 10^{-4}$; ALG8, $P = 2.59 \times 10^{-4}$; GANAB, $P = 5.52 \times 10^{-4}$; SEC61B, $P = 6.46 \times 10^{-4}$; ALG8, GANAB, and SEC61B all function in the common ER biogenesis pathway defined by SEC63 and PRKCSH. The increased occurrence of loss-of-function mutations and convergence on a single pathway known to be causative in PCLD prompted us to investigate these 3 genes as functional candidates for causing PCLD. In addition, the genome-wide significant enrichment for PKHD1 in our PCLD cohort prompted us to further explore the association of PKHD1 with PCLD.

Functional validation of candidates. We had previously shown that somatic homozygous inactivation of Prkcs and Sec63 in vivo results in polycystic kidney and liver disease due to defective biogenesis and trafficking of PC1 (7, 23, 33). This mechanistic understanding indicated that we can validate new candidate genes for PCLD by determining the effects of inactivation of each gene on alterations in its posttranslational maturation of PC1. Full-length PC1 (PC1-FL) undergoes autoproteolytic cleavage at the GPCR

### Table 2. Heterozygous mutations in ALG8, GANAB, SEC61B, and PKHD1 that result in PCLD

<table>
<thead>
<tr>
<th>Proband</th>
<th>Nucleotide change</th>
<th>Amino acid change</th>
<th>MAF in ExAC</th>
<th>Proband age</th>
<th>Sex</th>
<th>Kidney cysts</th>
<th>Liver cysts</th>
</tr>
</thead>
<tbody>
<tr>
<td>YU313</td>
<td>c.1090C&gt;T</td>
<td>p.R364X</td>
<td>4.12 × 10^{-5}</td>
<td>58</td>
<td>M</td>
<td>1</td>
<td>Innumerable, large</td>
</tr>
<tr>
<td>T-55</td>
<td>c.1090C&gt;T</td>
<td>p.R364X</td>
<td>4.12 × 10^{-5}</td>
<td>48</td>
<td>M</td>
<td>9</td>
<td>Innumerable, small</td>
</tr>
<tr>
<td>T-70</td>
<td>c.1090C&gt;T</td>
<td>p.R364X</td>
<td>4.12 × 10^{-5}</td>
<td>59</td>
<td>M</td>
<td>At least 3</td>
<td>Innumerable, small</td>
</tr>
<tr>
<td>FINN59</td>
<td>c.1038+1G&gt;T</td>
<td>Ex. 10 splice</td>
<td>1.65 × 10^{-4}</td>
<td>59</td>
<td>F</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>W-YU363</td>
<td>c.535C&gt;T</td>
<td>p.R175X</td>
<td>1.70 × 10^{-5}</td>
<td>63</td>
<td>M</td>
<td>3–4</td>
<td>Innumerable, many &gt;1 cm</td>
</tr>
<tr>
<td>T-116</td>
<td>c.2818C&gt;T</td>
<td>p.Q940X</td>
<td>Novel</td>
<td>47</td>
<td>F</td>
<td>At least 1</td>
<td>Innumerable, large</td>
</tr>
<tr>
<td>T9R62D5</td>
<td>c.1329delC</td>
<td>p.H441fs</td>
<td>Novel</td>
<td>50</td>
<td>M</td>
<td>At least 2, large</td>
<td>Innumerable small</td>
</tr>
<tr>
<td>T-87</td>
<td>c.152,153delCA</td>
<td>p.R51fs</td>
<td>8.24 × 10^{-6}</td>
<td>50</td>
<td>F</td>
<td>At least 10</td>
<td>Innumerable, small and large</td>
</tr>
<tr>
<td>YU356</td>
<td>c.102,103insAA</td>
<td>p.R34fs</td>
<td>Novel</td>
<td>42</td>
<td>F</td>
<td>1</td>
<td>Numerous, small</td>
</tr>
<tr>
<td>T-38</td>
<td>c.1A&gt;G</td>
<td>p.M1?</td>
<td>Novel</td>
<td>51</td>
<td>M</td>
<td>0</td>
<td>Innumerable, small</td>
</tr>
<tr>
<td>T-74</td>
<td>c.1242G&gt;T</td>
<td>p.Q404X</td>
<td>5.02 × 10^{-4}</td>
<td>64</td>
<td>F</td>
<td>At least 1</td>
<td>&gt;10, some &gt;1 cm</td>
</tr>
<tr>
<td>L-1224</td>
<td>c.12027C&gt;G</td>
<td>p.Y4009X</td>
<td>4.04 × 10^{-4}</td>
<td>–</td>
<td>F</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>T-61</td>
<td>c.1029G&gt;T</td>
<td>p.Q3407X</td>
<td>2.47 × 10^{-5}</td>
<td>55</td>
<td>M</td>
<td>1</td>
<td>Innumerable, small</td>
</tr>
<tr>
<td>T9R6321</td>
<td>c.9208C&gt;T</td>
<td>p.Q3070X</td>
<td>Novel</td>
<td>49</td>
<td>M</td>
<td>0</td>
<td>Innumerable, small</td>
</tr>
<tr>
<td>YU196</td>
<td>c.3766delC</td>
<td>p.Q1256fs</td>
<td>3.95 × 10^{-4}</td>
<td>–</td>
<td>F</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>T9R6216</td>
<td>c.3766delC</td>
<td>p.Q1256fs</td>
<td>3.95 × 10^{-4}</td>
<td>60</td>
<td>M</td>
<td>4 tiny cysts</td>
<td>Innumerable, small</td>
</tr>
<tr>
<td>T-60</td>
<td>c.1486C&gt;T</td>
<td>p.Q496X</td>
<td>5.68 × 10^{-4}</td>
<td>–</td>
<td>F</td>
<td>–</td>
<td>Innumerable, small</td>
</tr>
<tr>
<td>R2</td>
<td>c.880+1G&gt;T</td>
<td>Ex. 13 splice</td>
<td>Novel</td>
<td>–</td>
<td>F</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>T-58</td>
<td>c.820insA</td>
<td>p.R274fs</td>
<td>Novel</td>
<td>–</td>
<td>F</td>
<td>At least 1</td>
<td>Innumerable, small</td>
</tr>
<tr>
<td>T-59</td>
<td>c.107C&gt;T</td>
<td>p.T36M</td>
<td>5.20 × 10^{-4}</td>
<td>–</td>
<td>F</td>
<td>–</td>
<td>Innumerable, small</td>
</tr>
</tbody>
</table>

*All variants were found by whole exome sequencing and confirmed with Sanger sequencing. *Cyst counts are made from available reports or imaging. *African American. All other probands were European. *Affected first-degree relative (sibling or child) segregation data are available only for these 3 patients. Affected relatives also have the respective variant. *As this article was being prepared, the GANAB mutation in this patient’s sister, but not in this patient, was reported by Porath et al. (48).
HA epitope tags (Pkd1\(\text{\textit{F}}\)-BAC) (7, 33, 35). We used CRISPR/Cas9 genome editing methods to inactivate each candidate gene in the cells expressing Pkd1\(\text{\textit{F}}\)-BAC and investigated the impact on the PC1 protein.

ALG8. ALG8 (asparagine-linked glycosylation 8) encodes \(\alpha\)-1,3-glucosyltransferase, an ER integral membrane protein that functions to add the second of 3 glucoses to the assembling lipid-linked oligosaccharide precursor for N-linked glycosylation before it is transferred to the nascent peptide (38). ALG8 intersects with the protein biogenesis pathway between the sites of action of proteolysis site (GPS) into an approximately 3,000–amino acid extracellular N-terminal fragment (PC1-NTF) and an approximately 1,300–amino acid intramembranous C-terminal fragment (PC1-CTF) (34, 35). PC1-NTF migrates as 2 protein species — the cell surface–expressed endoglycosidase H–resistant (EndoH-resistant) PC1-NTR and the intracellular EndoH-sensitive PC1-NTS (36, 37). For the current study, we examined PC1 steady-state protein expression level and trafficking using a kidney epithelial cell line generated from mouse models containing 3 copies of a BAC transgene expressing Pkd1 with N-terminal FLAG and C-terminal HSP90 serves as loading control for cell lysates.

Figure 2. Mutations in ALG8 cause abnormal biogenesis of PC1. (A) YU313 has innumerable liver cysts larger than 1 cm (circle). (B) T-70 has innumerable small liver cysts (arrows). (C) W-YU363 has innumerable liver cysts, many larger than 1 cm (arrows, top panel), and 3–4 kidney cysts (arrows, bottom panel), while his 13-year-old daughter, YU364 (D), has the same ALG8 mutation but 8 kidney cysts (arrows, top and bottom panels) and no liver cysts. (E and F) Immunoblots of cell lysates with anti-HA (E) and anti-LRR PC1 N-terminal antibody (7e12) (F) show decreased PC1-CTF, PC1-FL, PC1-NTR, and PC1-NTS in Alg8\(\text{-/}\) cells. Re-expression of Alg8 in Alg8\(\text{-/}\) cells rescues PC1 expression (F). The labels in red mark the migration of the respective PC1 fragments in the Alg8\(\text{-/}\) cell lysate. (G and H) Anti-HA immunoprecipitation of PC1-FL and PC1-CTF was treated with PNGase F or EndoH. (G) Altered migration of PC1-FL is due to hypoglycosylation in Alg8\(\text{-/}\) cells. PNGase F treatment shows equivalent deglycosylated PC1-FL migration. (H) The relative proportion of EndoH-resistant/EndoH-sensitive PC1-CTF is reduced in Alg8\(\text{-/}\) cells. In G and H, the 2 PC1-CTF bands in the PNGase F–treated lanes result from alternatively spliced forms present in rodents (66). (I) Immunofluorescence shows absence of detectable PC1 (anti-HA) in cilia of Alg8\(\text{-/}\) cells. (J) Alg8\(\text{-/}\) cells activate the IRE1/\(\alpha\)/XBP1 branch of UPR, demonstrated by presence of XBP1s by reverse transcription PCR (RT-PCR; top) and immunoblotting (bottom). (K) Anti-HA immunoblot shows that inactivation of both Alg8 and Xbp1 does not change PC1 hypoglycosylation or GPS cleavage compared with Alg8 knockout alone.
SEC63 and GIIß. Patients with ALG8 mutations showed variation in severity of PCLD and frequent presence of kidney cysts (Figure 2, A–D, and Table 2). Four probands, YU313, T-70, T-55, and W-YU363, are male and have nonsense codon mutations in ALG8. Three of these probands, 2 European and 1 African American, shared the same mutation, p.R364X (Table 2). Identity-by-descent analysis found that the largest possible shared haplotype block among any 2 of these individuals is approximately 87 kb. This is too small to be indicative of a cryptic relationship between any of the probands and instead suggests that the mutation in each arose independently. FINN59 is a Finnish woman with a mild presentation carrying a splice mutation found with a frequency of 1.66 \times 10^{-3} in the Finnish population in ExAC. YU313 has symptomatic PCLD with many cysts larger than 1 cm (Figure 2A), while T-70 (Figure 2B) and T-55 (not shown) have numerous small liver cysts. W-YU363 has innumerable liver cysts of varied sizes (Figure 2C), but his daughter has 8 kidney cysts without any liver cysts (Figure 2D). She shares the ALG8 mutation with her father, and complete clinical testing for PKD1 and PKD2 mutations was negative; her mother has no kidney or liver cysts.

We used CRISPR/Cas9 to inactivate both ALG8 alleles in our mouse epithelial cell line expressing Pkd1F/H-BAC. Sanger sequencing confirmed compound heterozygous frameshifting loss-of-function mutations in ALG8-null cells (ALG8+/–; Supplemental Table 6). ALG8+/– cells showed reduced levels of PC1-FL and PC1-CTF detected by the anti-HA C-terminal epitope tag (Figure 2E). PC1-NTR and PC1-NTS detected by the anti-N-terminal native protein antibody were also reduced (Figure 2F). The 2 antibodies are detecting different pools of PC1 in these cells. The anti-HA antibody is detecting PC1 expressed from the Pkd1F/H-BAC transgene, whereas the N-terminal anti-LRR native PC1 antibody (7e12) is primarily detecting Pkd1 expressed from the endogenous gene (Supplemental Figure 2). The presence of the FLAG epitope in the transgenically expressed PC1 appears to impair recognition of the native epitope. There was no difference in Pkd1 transcript expression between WT and ALG8+/– cells (data not shown), indicating that differences in steady-state protein level are a post-transcriptional effect. Examination of integral membrane proteins from other membrane compartments, i.e., PC2 (ER, cilia), Na+/K+-ATPase \( \alpha \) subunit (plasma membrane), and calnexin (ER), did not show significant quantitative reductions, indicating that not all membrane proteins are similarly affected by inactivation of ALG8 (Supplemental Figure 3A). Since CRISPR/Cas9 can produce off-target effects, we showed the specific dependence of these findings on ALG8 by rescuing the protein expression level of PC1 with re-expression of ALG8 cDNA in ALG8+/– cells (Figure 2F).

The electrophoretic migration of PC1-FL was notably faster in ALG8+/– cells (Figure 2, E–G and K), suggesting that PC1 may have a defect in glycosylation in addition to reduced protein quantity in the absence of ALG8. Immunoprecipitation of PC1-FL and PC1-CTF followed by treatment with peptide-N-glycosidase F (PNGase F) resulted in the same electrophoretic migration for the protein backbones irrespective of ALG8 genotype, confirming that differences in glycosylation account for the genotype-dependent altered gel mobility of PC1-FL in cell lysates (Figure 2G). The data show that effective reduction in PC1 function in the absence of ALG8 is likely to result from the combined effect of the quantitative decrease in steady-state levels of PC1 and the structural abnormality of PC1 resulting from defective posttranslational modification. N-glycosylated proteins typically acquire resistance to EndoH as they transit past the middle Golgi on the way to the cell surface. The relative proportion of EndoH-resistant PC1-CTF to EndoH-sensitive PC1-CTF is markedly reduced in ALG8+/– cells (Figure 2H), providing biochemical evidence that PC1 trafficking to the cell surface, and by extension to cilia, is markedly impaired following inactivation of ALG8. In keeping with this, immunofluorescent cell staining with anti-ARL13B to mark cilia and anti-HA to detect PC1 shows absent or markedly reduced visualization of PC1 in cilia in ALG8+/– cells compared with cells expressing ALG8 normally (Figure 2I).

We had previously shown that inactivation of Sec63, but not of Prkcsb, results in activation of the IRE1a/XBP1 branch of the unfolded protein response (UPR) (33). Activation of this branch of UPR results in formation of XBP1s, the active spliced form of XBP1, which acts as a transcription factor for an array of chaperone proteins (39). ALG8+/– cells also have activation of the IRE1a/XBP1 branch of UPR as evidenced by reverse transcription PCR and immunoblot assays showing the presence of XBP1s (Figure 2J). In the absence of SEC63, active XBP1s is required to enable autoproteolytic cleavage of PC1 at the GPS (33). To determine whether XBP1s has a similar role in ALG8+/– cells, we inactivated XBP1 in ALG8+/– cells to make double-knockout cells (ALG8+/– Xbp1+/–; Figure 2K). ALG8+/– Xbp1+/– cells showed normal cleavage of PC1 (Figure 2K), indicating that while the IRE1a/XBP1 branch of UPR is activated following loss of ALG8, it is not required for PC1 cleavage. In aggregate, the relative enrichment of rare loss-of-function variants of ALG8 in PCLD patients coupled with functional evidence that inactivation of ALG8 affects PC1 biogenesis and trafficking supports the conclusion that ALG8 mutations are a genetic cause of PCLD.

GANAB. Three PCLD patients, but none in the control population, had loss-of-function mutations in GANAB encoding GIß (Tables 1 and 2). The severity of the liver phenotype in GANAB patients varied from numerous large liver cysts with hepatomegaly (Figure 3A) to innumerable small cysts (Figure 3B). All 3 patients had at least 1 kidney cyst, but only 1 had more than 10 (Figure 3, A and B, and Table 2). To assess the functional consequences of loss of GANAB, we inactivated the gene in the cells expressing Pkd1F/H-BAC using CRISPR/Cas9. We confirmed inactivation of GANAB by Sanger sequencing that showed homozygous deletion of 281 bp beginning in exon 12 and extending into intron 13 (Supplemental Table 6) and by anti-GANAB immunoblot of cell lysates (Figure 3C). GANAB+/– cells showed markedly reduced or absent PC1-NTR (Figure 3C) and decreased steady-state levels of PC1-FL and PC1-CTF (Figure 3D). The Na+/K+-ATPase \( \alpha \) subunit protein was also decreased in the absence of GANAB, suggesting effects on at least some other plasma membrane proteins (Supplemental Figure 3, B and C). Functional re-expression of GIß in GANAB+/– cells rescued the expression of PC1-NTR, showing that the observed effects on PC1 protein maturation were solely the result of GIß deficiency (Figure 3C).

A reproducible \((n = 4)\) feature of GANAB+/– cells was that PC1-FL and PC1-NTS migrated more slowly (Figure 3, C and D). We
had previously shown that there is aberrant glucose trimming with GIIβ mutations (40), so the slower electrophoretic migration of PC1 fragments in Ganab−/− cells. Immunoblot of cell lysates with anti–PC1-LRR N-terminal antibody (C, top panel) and anti-GANAB (C, GIIα; middle panel) demonstrates absence of PC1-NTR (asterisk) in Ganab−/− cells. Anti-HA immunoblot (D) shows significant decrease in PC1-CTF. PC1-FL (C and D) and PC1-NTS (C) have slower migration in Ganab−/− cells. The PC1 expression level and altered gel-migration pattern are rescued by re-expression of Ganab (C, Ganab−/−;Ganab+). The re-expressed GIIα is larger than the native protein because of the inclusion of an epitope tag and an alternatively spliced exon 6 that is absent from the native protein in the mouse cell line. (E) PC2 and PC1-FL show higher molecular mass in Ganab−/− cells that resolves to the same gel migration in WT cells following PNGase F treatment. *Nonspecific band. (F) Ganab−/− cells only have EndoH-sensitive (S) and lack EndoH-resistant (R) PC1-CTF. PC1-CTF shows 2 bands due to alternative splice forms in rodents. (G) IRE1α/XBP1 branch of UPR is activated in Ganab−/− cells as evidenced by increased XBP1s. Tunicamycin (Tun) treatment (2.5 μg/ml for 6 hours) serves as positive control for UPR activation. (H) Functional assay of GANAB missense variants. Fluorescence output from the glucosidase II fluorogenic substrate 4-methylumbelliferyl α-D-glucopyranoside (4-MUG) from HEK 293T cells cotransfected with GIIβ and the respective variants of GIIα. Background fluorescence from untransfected cell lysates is subtracted. HSP90 serves as loading control for cell lysates.

The residual protein is abnormally processed and has markedly impaired trafficking past the Golgi. While there is no activation of the IRE1α/XBP1 branch of UPR by inactivation of GIIβ (33), we found that loss of GIIα resulted in increased formation of the active XBP1s, suggesting a somewhat more severe cellular stress with inactivation of the catalytic subunit of GII (Figure 3G).

Two patients in our cohort had rare, potentially deleterious GANAB missense variants p.H807N and p.H872Y (Supplemental Table 7). Each occurred in individuals in whom a causative PCLD gene mutation in ALG8 and PKHD1, respectively, had been identified. We expected that these GANAB variants do not result in complete loss of function and are therefore not causative for PCLD. To test this hypothesis, we assayed the catalytic activity of both GIIα variants and compared them with WT GIIα and with a known loss-of-function mutant we had identified, p.R51fs. HEK 293T cells cotransfected with GIIβ and the respective variants of GIIα. Background fluorescence from untransfected cell lysates is subtracted. HSP90 serves as loading control for cell lysates.
confirmed complete loss of function by both direct sequencing of cell genomic DNA (Supplemental Table 6) and immunoblotting with native anti-SEC61β antisera (Figure 4B). Inactivation of SEC61β resulted in almost undetectable levels of PC1 when examined by either C-terminal anti-HA epitope antibody or N-terminal native protein antibody (Figure 4B). Very long exposure of the latter immunoblot showed a minute amount of PC1-FL and PC1-NTS, but no PC1-NTR was detectable in Sec61b–/– cells (Figure 4B).

In some experiments, a long exposure of a C-terminal anti-HA immunoblot could detect minute amounts of PC1-FL, but PC1-CTF remained completely undetectable (Figure 4C). As with the other PCLD genes, inactivation of SEC61β had variable effects on other integral membrane proteins — Na+/K+-ATPase α1 subunit was markedly reduced, PC2 was minimally affected, and calnexin was unchanged (Supplemental Figure 3, C–E). Re-expression of SEC61β in Sec61b–/– cells reconstituted PC1 levels including PC1-CTF, confirming the specificity of this effect (Figure 4D). As an additional indication that SEC61β inactivation has severe effects, lysates were incubated with the GII fluorogenic substrate 4-methylumbelliferyl α-D-glucopyranoside (4-MUG), and fluorescence output above the baseline values obtained using untransfected cell lysates was used to quantify GII enzymatic activity (Figure 3H and ref. 40). The activity of p.H872Y was similar to WT, while p.H807N showed approximately 50% reduction in activity (Figure 3H). The pathogenic variant p.R51fs showed complete loss of activity. While these missense variants in GIIα are not pathogenic in PCLD, it remains possible that heterozygous hypomorphic variants such as p.H807N could have genetic modifier effects on the expressivity of PCLD due to mutations in other causative genes (e.g., PKHD1).

SEC61β. The SEC61 protein translocation pore consists of 3 subunits, α, β, and γ. The β subunit, encoded by SEC61B, is the least conserved component and the one that is not essential for SEC61 complex function (32). Two patients had heterozygous loss-of-function variants in SEC61B (Tables 1 and 2). In both patients, PCLD is characterized by diffuse and numerous small liver cysts (Figure 4A). We inactivated SEC61β in the Pkd1F/H-BAC cells and confirmed complete loss of function by both direct sequencing of cell genomic DNA (Supplemental Table 6) and immunoblotting with native anti-SEC61β antisera (Figure 4B). Inactivation of SEC61β resulted in almost undetectable levels of PC1 when examined by either C-terminal anti-HA epitope antibody or N-terminal native protein antibody (Figure 4B). Very long exposure of the latter immunoblot showed a minute amount of PC1-FL and PC1-NTS, but no PC1-NTR was detectable in Sec61b–/– cells (Figure 4B). In some experiments, a long exposure of a C-terminal anti-HA immunoblot could detect minute amounts of PC1-FL, but PC1-CTF remained completely undetectable (Figure 4C). As with the other PCLD genes, inactivation of SEC61β had variable effects on other integral membrane proteins — Na+/K+-ATPase α1 subunit was markedly reduced, PC2 was minimally affected, and calnexin was unchanged (Supplemental Figure 3, C–E). Re-expression of SEC61β in Sec61b–/– cells reconstituted PC1 levels including PC1-CTF, confirming the specificity of this effect (Figure 4D). As an additional indication that SEC61β inactivation has severe effects...
The Journal of Clinical Investigation

RESEARCH ARTICLE

on cellular protein biogenesis, Sec61β−/− cells showed the strongest relative activation of XBP1s among all of the PCLD gene cellular knockouts (Figure 4E). In aggregate, the data show that loss of SEC61β results in the most severe quantitative reduction of PC1 among the PCLD genes.

PKHD1. Patients were ascertained for this study based solely on imaging criteria defining PCLD. Whole exome analysis of the discovery cohort identified heterozygous loss-of-function variants in PKHD1, the ARPKD gene, in 9 individuals (Tables 1 and 2). A tenth individual had the p.T36M missense variant in PKHD1 (Table 2), which is a known pathogenic founder mutation in Europeans and the most common mutant allele (17%) in ARPKD (41). Therefore, a minimum of 10 of 102 PCLD probands in our discovery cohort are explained by mutations in PKHD1. This significant enrichment of heterozygous pathogenic PKHD1 variants in the PCLD cohort relative to the observed occurrence of these mutations in the general population achieved genome-wide significance (Figure 1 and Tables 1 and 2). The clinical presentation of most of these patients showed innumerable small liver cysts (Figure 5, A–E, and Table 2). This is in contrast to the more typical cyst architecture of PCLD in patients with SEC63, PRKCSH, ALG8, or GANAB variants, which often includes multiple large cysts greater than 1 cm in size. This may indicate dependence of the clinical phenotype of the PCLD presentation...
Gene mutations had somewhat rare (<5 × 10⁻³ in ExAC) or novel assays. Seven probands not explained by any of the other PCLD is difficult to assess in the absence of robust functional PKHD1 action (T-74) had macrocysts in the liver and 1 kidney cyst (Figure 6). Mutations on the genic etiology, although at least 1 case with a PKHD1 mutation (T-74) had macrocysts in the liver and 1 kidney cyst (Figure 5F). The pathogenicity of other uncommon missense variants in PKHD1 is difficult to assess in the absence of robust functional assays. Seven probands not explained by any of the other PCLD gene mutations had somewhat rare (<5 × 10⁻³ in ExAC) or novel missense variants in PKHD1 with combined annotation dependent depletion (CADD) scores greater than 14 (Supplemental Table 7). Four of the seven missense variants are predicted by MetaSVM (42), another aggregate prediction algorithm, to be deleterious (Supplemental Table 7). No comparably rare PKHD1 missense variants were found in any patient with mutations in the other known causative genes (SEC63, PRKCSH, ALG8, GANAB, or SEC61B). In aggregate, mutations in PKHD1 explain at least approximately 6% (10/159), and may explain up to approximately 11% (17/159), of clinically ascertained PCLD probands. While we did not find loss-of-function mutations nor any of the published pathogenic missense variants in LRPS in our cohort, we did find 4 missense variants with MAF less than 1 × 10⁻³ and CADD greater than 15 (Supplemental Table 8). All 4 of these variants occurred in cases that either were explained by loss-of-function mutations in PKHD1 (T-58 and T-59; Table 2) or may be explained by missense variants in PKHD1 (TOR6467 and TOR2400; Supplemental Tables 7 and 8). This does raise the possibility that genetic interactions with genes such as LRPS5 may be a factor in determining the penetrance of PCLD in PKHD1 carriers.

PKHD1 is not known to belong to the ER protein biogenesis pathway defined by the other PCLD genes. We addressed this experimentally by inactivating Pkd1 in Pkd1⁺/⁻-BAC cells. We produced 2 independent cell lines with either homozygous or compound heterozygous frameshift mutations in exon 3 of Pkd1 (Supplemental Table 6). Neither Pkd1-null (Pkd1⁻/⁻) cell line showed any changes in PCI-FL, PCI-CTF, PCI-NTR, or PCI-NTS levels when examined by anti-HA epitope or native protein N-terminal antiseras (Figure 5G). There was also no change in quantity of PC2 (Figure 5G). In keeping with the biochemical data, PCI was readily detected in cilia of Pkd1⁻/⁻ cells (Figure 5H). The evidence suggests that PKHD1 acts by a different mechanism than the protein biogenetic pathway impacted by the other PCLD genes. In aggregate, this study defines causative PCLD mutations in 4 genes that account for at least 20 probands in our discovery cohort of 102 and suggests that at least 2 different pathways are potential causes of clinically defined PCLD.

**Discussion**

Dominantly inherited mutations in SEC63 and PRKCSH cause PCLD by reducing the effective functioning dosage of PCI in cells that undergo somatic second-hit mutations and result in recessive loss of function of the respective PCLD gene (7, 23). Once PCI activity falls below a threshold, the rate of cyst progression is directly related to the degree of impairment of PCI activity. This continuum also applies to ADPKD (43, 44), but bile ducts appear to have a lower tolerance for reduced PCI dosage compared with kidney tubules. This can be due to either a higher threshold requirement or a lower baseline level of PCI activity in bile duct cells, which may explain the occurrence of PCLD rather than ADPKD with genetic defects affecting PCI biogenesis. Identification of additional genes that impair PCI functioning and manifest as human PCLD offers the opportunity to deepen our understanding of both ADPKD and PCLD. The advent of whole exome sequencing has enabled gene discovery that is not dependent on family data and is tolerant of diseases, such as PCLD, with extensive genetic heterogeneity (28–30). Applying this approach to studying the approximately 65% of clinically ascertained PCLD probands without mutations in SEC63 or PRKCSH, we found that no single gene explains more than 11% of the unknown cases. This suggests that there may be upwards of 10 to 15 causative genes for PCLD.
Three genes identified in the current study, ALG8, GANAB, and SEC61B, fit into the mechanistic paradigm defined by SEC63 and PRKCSH. They encode proteins that function in the post-translational protein biosynthetic pathways in the ER for integral membrane and secreted proteins (Figure 6). Analysis of in vitro cell models with knockout of each gene shows reduction in the steady-state levels of PC1 as a consistent feature. Beyond this common theme, there is substantial variation in the effects of each gene on PC1. Loss of Sec61b results in a substantially more severe deficiency of PC1 than loss of any of the other PCLD genes, including Sec63 and Prkcs. Alg8 knockout causes hypoglycosylation of PC1, whereas Ganab knockout results in defective glucose trimming of N-glycan moieties that produces an increased apparent molecular mass for PC1. Inactivation of each of these 3 genes, including Ganab, results in activation of at least the Ire1s/Xbp1 branch of UPR. This differs from Prkcs, which does not activate UPR (33) despite the fact that the respective protein products of Prkcs and Ganab work together as subunits of the G1I holoenzyme. This suggests that the “cellular severity” of loss of the catalytic subunit, Ganab, is greater than that of the loss of the regulatory subunit, Prkcs, required for ER retention. Whereas activation of Xbpls in Sec63 knockouts is required to support GPS cleavage of PC1 (33), we found that Xbpls does not play this role in cells lacking Alg8.

Our finding that the pathogenic mechanisms of 5 PCLD genes converge on posttranslational modulation of PC1 function makes plausible the extended hypothesis that all genetic causes of PCLD occur by indirect impairment of effective PC1 signaling. This does not imply that all other PCLD genes affect PC1 posttranslational processing. Genetic defects that cause significant loss of PC1 activity by other indirect mechanisms may also result in PCLD. For example, PC1 acts as a negative regulator of an unidentified cilia-dependent signal that promotes cyst formation when intact cilia lack PC1 (45). Loss of function of any factor that works with PC1 in cilia as part of this negative regulatory activity may cause PCLD by impairing the efficiency of PC1 action.

PKHD1 and LRPS are 2 examples of genes not clearly associated with PC1 biogenesis that are nonetheless associated with PCLD. We found genome-wide significant enrichment of heterozygous PKHD1 loss-of-function carriers in our cohort of clinically ascertained PCLD. This genetic discovery is consistent with a clinical report on the occurrence of innumerable asymptomatic liver cysts in approximately 10% of ARPKD carrier parents (46). Whether reduced or absent fibrocytin causes PCLD due to impaired PC1 signaling, or whether it causes PCLD independently of PC1, cannot be definitively answered with the available data. The former possibility is supported by strong evidence for a genetic interaction between Pkd1 and Pkd1 in mouse models. Reduced Pkd1 dosage exacerbates the kidney and liver cysts in mice with homozygous Pkd1 mutations (7, 24). This raises the possibility that somatic inactivation of PKHD1 may have a postdevelopmental role in bile ducts that sufficiently impairs effective signaling by PC1 in penetrant individuals to manifest with the sporadic cysts seen in PCLD.

LRPS, a coreceptor for Fizzled in canonical Wnt signaling, has been proposed as a causal gene for PCLD (18). A recent study proposed that PC1 is a coreceptor for noncanonical Wnt signaling (47). These data may suggest an indirect functional relationship between LRPS and PC1. We did not find any loss-of-function mutations in LRPS, nor could we attribute any of the cases in our whole-exome-sequenced cohort to mutations in LRPS. However, we did note co-occurrence of indeterminate LRPS missense variants with PKHD1 variants. The phenotypes in these individuals did not stand out, but the co-occurrence of LRPS and PKD1 variants raises the possibility that incompletely penetrant heterozygous variants in LRPS, and by extension other genes, may play a role in determining the penetrance of the PCLD phenotype in PKHD1 carriers.

Since the absolute number of PCLD patients with mutations in PKHD1 is small, it is not possible to draw definitive conclusions about clinical presentations. Nonetheless, some suggestive observations are worth noting. Five of the six PCLD patients explained by PKHD1 for whom imaging was available presented with a multitude of small cysts. This was also characteristic of the study that ascertained carrier parents of ARPKD children (46). This is in contrast to the larger cysts typical of most of the other PCLD genes and ADPKD. It fibrocytin is working through a PC1-dependent mechanism, a plausible hypothesis regarding this presentation is that the relative impairment of PC1 function due to loss of fibrocytin is less than that observed with loss of the other PCLD genes. Interestingly, the 2 cases of PCLD due to SEC61B mutations also presented with small cysts, yet SEC61B appears to have the most profound quantitative reduction in PC1 of all of the PCLD genes. This apparent paradox with the “PC1 dosage” hypothesis may be explained by the more severe and global effect of SEC61B inactivation on the rest of its non-PC1 client proteins. The drive toward tissue remodeling and cyst growth due to almost absent PC1 in bile duct epithelial cells lacking SEC61B may be partly counteracted by poor overall biosynthetic efficiency of other proteins required for cells to grow larger cysts. In future studies, clinically stratifying patients into those with macrocystic PCLD and those with consistently small cysts detected by MRI may enrich subsets of causative genes in these clinical subgroups.

ALG8 presented the most pleiotropic clinical manifestations of PCLD and highlights that PCLD cyst growth may also be modified by additional genetic or nongenetic factors beyond the established association with female sex (2, 3). One male patient with ALG8 mutation had severe PCLD that required procedural intervention to reduce liver cyst mass, while another had only minute cysts more typical of PKHD1 and SEC61B. Perhaps most interestingly, in a familial case attributed to ALG8, the father had typical PCLD that included a few kidney cysts, while his daughter had 8 kidney cysts and no liver cysts at age 19. This suggests that ALG8 may cross over clinically between PCLD and ADPKD. While this work was being prepared for publication, GANAB was independently identified as causing liver and kidney cysts in a population with a significantly greater burden of kidney cysts who were presumed to have ADPKD but who were excluded for mutations in PKD1 and PKD2 (48). The 2 novel GANAB patients in our study also have kidney cysts. Overall, it appears that GANAB and ALG8 may straddle the diagnoses of PCLD and ADPKD, and mutations in ALG8 may also explain some cases of apparent ADPKD that do not have mutations in PKD1 or PKD2 or GANAB. One mechanistic hypothesis for this multigene phenotype may be that GANAB and ALG8 seve-
ly affect PCI1 maturation, but without the more global protein biosynthetic defects hypothesized for SEC61B that counter the cellular responses required for cyst growth.

PCLD due to ALG8, like that due to PKHD1, occurs in adult carriers of a recessive disorder. ALG8 is the only one of the ER-associated PCLD genes for which a recessive genotype is reported to be compatible with life. Fifteen cases of congenital disorder of glycosylation 1H (CDG1H) have been reported in the literature, caused by homozygous or compound heterozygous combinations of loss-of-function and deleterious missense mutations in ALG8 (49). The CDG1H phenotype has been reported to include renal failure and hepatomegaly with at least 1 report of bile duct dilation and bile duct cysts (50). Other features include protein-losing enteropathy, hypoalbuminemia, edema, ascites, and death in early infancy or childhood (51). Screening of ALG8 heterozygous carrier parents by imaging may show an increased occurrence of PCLD much as was found in ARPKD carrier parents (46). This is the first instance of data supporting the hypothesis that PCLD may manifest in carriers of severe recessive diseases, particularly congenital disorders of glycosylation (50). Notably, recessive mutations in ALG9 cause CDG1L and are associated with polycystic kidneys (52, 53). Heterozygous carriers of ALG9 mutations may manifest PCLD as well.

With the identification of 4 additional disease genes that can cause the PCLD phenotype, we have a genetic explanation of almost 50% of cases in our cohort of 159 probands. The newly identified ER genes confirm our previous discovery that defective PC1 biogenesis in the ER is central to the PCLD phenotype. This knowledge, the previous data that Pkd1 and Pkhd1 interact genetically, and our current finding that PKHDI and PCLD genes provide increasing support for the hypothesis that the PCI and fibrocytin signaling pathways intersect and each is sensitive to reduced dosage of the other. The co-occurrence of mild kidney cysts in some cases of PCLD supports the hypothesis that PCLD and ADPKD represent a clinical continuum that is partly a function of the severity of impairment of PCI1 by the different genetic causes. The convergence on PCI1 dependence of these phenotypes sets the dominant polycystic diseases apart from the recessive ciliopathy spectrum of disease. Discovery of another novel PCLD gene will identify additional potential modifiers of PCI function. These genes may encode proteins functioning beyond the ER, perhaps even in cilia. These may prove to be powerful clues to unraveling the still-elusive function of the polycystin proteins.

Methods

Cohort recruitment and exome analysis. Index cases meeting entry criteria (see Results) for PCLD and at-risk family members were enrolled by study investigators at Yale, or sent to us deidentified by coinvestigators at other sites (2, 12, 13). Representative liver and kidney imaging was obtained from the referring physician when available. Genomic DNA was obtained from the referring physician or extracted from peripheral blood leukocytes using Qiagen Puregene kit reagents. Genomic DNA from the proband was sent for targeted exome capture using the NimbleGen/Roche SeqCap EZ Exome v2 reagent, or MedExome reagent, followed by 75-px paired-end sequencing using Illumina HiSeq 2000. Sequence reads were aligned to the GRCh37/hg19 human reference genome using the Burrows Wheeler Aligner algorithm BWA-MEM (54). Variants were called using the Genome Analysis Toolkit Haplotype Caller (55). Annotations were carried out using scripts designed by the authors and members of the Yale Center for Genome Analysis. The analysis pipeline has been validated (29, 56) and makes use of SAMtools and Genome Analysis Toolkit utilities (55), and ANNOVAR software (57).

Bioinformatics. The observed gene burden of rare (ExAC frequency <1 × 10−5) loss-of-function alleles (nonsense, frameshift, splice site) was compared with the expected burden independent of ethnicity, and that in the European subset of discovery cohort cases (n = 92) was compared with that in European controls (n = 3,274) using existing scripts and statistical tests listed below. The European controls were composed of healthy parents of probands with autism and congenital heart disease sequenced using the same capture reagent, platform, and analysis pipeline as study cases. The Q-Q plot was made using R software (https://www.r-project.org). PCA analysis used the EIGENSTRAT program (58). Pairwise kinship analysis was performed using PLINK (59). Identity-by-descent (IBD) analysis was performed using the phasing algorithm and IBD detection algorithm in the Beagle 4.1 software package (60, 61).

Statistics. P values for observed versus expected gene burden comparisons were calculated using the binomial test, and for case versus control analysis using the Fisher’s exact test. Statistical assessment of immuno blot densitometry data was done using paired-sample 2-tailed t test.

In vitro gene knockout. All in vitro gene knockout studies were done in an immortalized mouse renal proximal tubule epithelial cell line containing 3 copies of a Pkd1−/− bacterial artificial chromosome (BAC) transgene (7). CRISPR/Cas9 was used to generate clonal cell lines containing protein-truncating mutations in the respective genes. Guide RNA sequences (Supplemental Table 6) were designed using the CRISPR Design tool (http://crispr.mit.edu/) (62) for Alg8, Xbp1, Pkh1, Pkd1, and Sec61b, and using CRISPRScan (http://www.crisprscan.org) (63) for Ganab. For Alg8, Pkd1, and Xbp1 knockouts, cells were infected with lentivirus containing lentiCRISPR v2 (Addgene plasmid 52961) (64). Following selection, clones were separated by clonal dilution, and screened by Sanger sequencing for frameshifting mutations. Ganab−/− and Sec61b−/− cell lines were generated in a population of WT cells stably expressing Cas9 from lentiCas9-Blunt (Addgene 52962). For each gene targeted, a large deletion was made between the sites of 2 guide RNAs in independent modified-pGL3-U6-sgRNA-PGK-puro vector (Addgene 51133) expression plasmids transiently transfected by electroporation. Clones were screened with PCR, then verified by Sanger sequencing. Re-expression of Alg8, Ganab, and Sec61b cDNA was done with lentiviral infection using the pCDH-CMV-MCS-EF1-Puro vector (System Biosciences) with a modified resistance cassette.

Immunoblotting. Cells were harvested and lysed in RIPA buffer containing Complete Protease Inhibitor Cocktail (Roche). Lysate was run on Novex 3%–8% Tris-acetate protein gels or Bio-Rad 10% Tris-glycine protein gels, and transferred to PVDF membranes. Immunoblotting was performed using the primary antibodies listed below, and detected using Licor Odyssey Fc imaging system after application of HRP-conjugated secondary antibodies and washing with triethanolamine-buffered saline with Tween-20 (TBST).

Primary antibodies. The primary antibodies used were anti-HA (3F10, Roche; dilution 1:3,000), anti-PC1-NTF (7e12, Santa Cruz Biotechnology), anti-PC2 (YCC2, S. Somlo; 1:20,000), anti-HSP90 (BAC) transgene (7). CRISPR/Cas9 was used to generate clonal cell lines containing 3 copies of a Pkd1−/− bacterial artificial chromosome (BAC) transgene (7). CRISPR/Cas9 was used to generate clonal cell lines containing protein-truncating mutations in the respective genes. Guide RNA sequences (Supplemental Table 6) were designed using the CRISPR Design tool (http://crispr.mit.edu/) (62) for Alg8, Xbp1, Pkh1, Pkd1, and Sec61b, and using CRISPRScan (http://www.crisprscan.org) (63) for Ganab. For Alg8, Pkd1, and Xbp1 knockouts, cells were infected with lentivirus containing lentiCRISPR v2 (Addgene plasmid 52961) (64). Following selection, clones were separated by clonal dilution, and screened by Sanger sequencing for frameshifting mutations. Ganab−/− and Sec61b−/− cell lines were generated in a population of WT cells stably expressing Cas9 from lentiCas9-Blunt (Addgene 52962). For each gene targeted, a large deletion was made between the sites of 2 guide RNAs in independent modified-pGL3-U6-sgRNA-PGK-purovector (Addgene 51133) expression plasmids transiently transfected by electroporation. Clones were screened with PCR, then verified by Sanger sequencing. Re-expression of Alg8, Ganab, and Sec61b cDNA was done with lentiviral infection using the pCDH-CMV-MCS-EF1-Puro vector (System Biosciences) with a modified resistance cassette.
(H114, Santa Cruz Biotechnology; 1:10,000), anti-GAPDH (14C10, Cell Signaling), anti-calnexin (ADI-SPA-860, Enzo), anti-α-Na+/K+-ATPase raised against chicken α-Na+/K+-ATPase amino acids 338–724 (a gift of M.J. Caplan, Yale University, New Haven, Connecticut, USA) (65), anti-XBP1 (Poly6195, Biolegend; recognizes the C-terminal 15 amino acids of XBP1), rabbit polyclonal anti-SEC61B (gift from R. Zimmerman, University of Saarland, Hamburg, Germany), anti-MYC (9e10, Santa Cruz Biotechnology), and anti-α-glucosidase II (ab179805, Abcam; 1:2,000). All antibodies were used at 1:1,000 dilution in 5% milk in TBST unless otherwise indicated.

Investigation of glycosylation status. Cell lysate was obtained in RIPA lysis buffer and incubated with anti-HA–conjugated agarose beads (Sigma-Aldrich). After washing, protein-bound beads were treated with EndoH or PNGase F (New England Biolabs). Treated PC1-HA protein was eluted from the beads by heating in 2X sample buffer (Sigma-Aldrich).

Immunofluorescence. Cells were cultured on glass cover slips to confluence and allowed to differentiate at 37°C without insulin/transferrin/selenium supplement (ITS), T3, or IFN-γ for 10–14 days (7). They were serum-starved for 2 days to form cilia before analysis. Cells were fixed with 4% paraformaldehyde and permeabilized with 0.2% Triton before blocking (0.1% BSA plus 10% goat serum) and application of primary antibodies. After washing and fluorescent secondary antibody application, slides were imaged by confocal microscopy (Nikon) and analyzed using Elements software (Nikon).

Quantitative PCR and reverse transcription PCR. RNA was isolated from cultured cells using Trizol Reagent (Invitrogen). cDNA was reverse transcribed from RNA using reagents from New England Biolabs. Primers for Pkd1 (forward, ACACGCTTGGGAATGGA; reverse, CCATGGGAGATGGTCTGGG) were designed using NCBI’s primer designing tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). Quantitative PCR was done by Bio-Rad CFX Connect Real-Time PCR Detection System.

GII enzymatic testing. pCDNA3.1 eukaryotic expression plasmid expressing human GANAB cDNA sequence (gift of Anne Marcil, National Research Council, Ottawa, Ontario, Canada) was edited using site-directed mutagenesis methods. HEK 293T cells were cotransfected with pCDNA3.1 containing PRKCSH cDNA sequence and pCDNA3.1 containing GANAB sequence using Lipofectamine 2000 (Invitrogen). Lysate was made 72 hours after transfection using RIPA lysis buffer. Lysate was then incubated with a GII substrate, 1 mM 4-methylumbelliferyl α-D-glucopyranoside (4-MUG, Toronto Research Chemicals), at 37°C for 30 minutes, and reaction was continued for 2 M Tris. Fluorescence was measured with excitation/emission wavelengths 365/445 nm. Each transfection was done in triplicate, and each sample was tested with substrate in triplicate. Fluorescence values from empty vector–transfected cell lysate were subtracted from those of experimental samples to remove background native GII activity.

Study approval. This study was approved by the institutional review boards at both the Yale University School of Medicine and the Mayo Clinic. Written informed consent was obtained from all study participants prior to enrollment, and all samples and results were described and stored using a deidentified study code.

Author contributions. WB organized the cohort, carried out bioinformatics, designed and conducted experiments, and drafted the manuscript. KD and SVF developed cell lines and optimized experimental protocols. JC and MC conducted bioinformatics analysis. SP identified ALG8 as a candidate gene and identified PKHD1 variants. ARG provided reagents and experimental design for PKHD1 studies. JK and MC designed exome analysis pipelines and annotation scripts. EBH and AG performed Sanger sequencing verification of mutations. ET, PT, TW, YPP, and VET contributed patient samples and clinical information. SSC provided family studies. SM and RPL provided design of exome sequencing strategies. RPL designed bioinformatic evaluation. SS conceived and designed the project, organized collaborators, designed experiments, and edited the manuscript. Funding was obtained by WB and SS.

Acknowledgments. We gratefully acknowledge the generous participation of all the patients and families in this study. We thank Neera Dahl, Tamar Taddei, Simona Jakab, Joan Cho, James Boyer, Luca Fabris, Reinier Wiest, and David Roer for referring patients with PCLD. We thank Ali G. Gharavi for thoughtful discussions and critical reading of the manuscript, and Arnaud Marlier for contributing artwork. This study was supported by NIH grants R01 DK51041 and R01 DK54053 to SS; T32 DK007276 and a Polycystic Kidney Disease Foundation Fellowship to WB; the Yale Center for Mendelian Genomics (NIH U54 HG006504, to RPL); the Mayo Clinic PKD Center (P30 DK090728, to VET); and the Yale O’Brien Kidney Center (P30 DK079310).

Address correspondence to: Stefan Somlo, Section of Nephrology, Yale University School of Medicine, PO Box 208029, 333 Cedar Street, New Haven, Connecticut 06520-8029, USA. Phone: 203.737.2974; E-mail: stefan.somlo@yale.edu.


