**ALG9 Mutation Carriers Develop Kidney and Liver Cysts**

Whitney Besse,1 Alex R. Chang,2 Jonathan Z. Luo,3 William J. Triffo,4 Bryn S. Moore,3 Ashima Gulati,1 Dustin N. Hartzel,5 Shrikant Mane,6 Regeneron Genetics Center, Vicente E. Torres,7 Stefan Somlo,1,6 and Tooraj Mirshahi5

Departments of 1Internal Medicine (Nephrology) and 6Genetics, Yale University School of Medicine, New Haven, Connecticut; Departments of 2Nephrology, 3Molecular and Functional Genomics, 4Radiology, and 5Biomedical and Translational Informatics, Geisinger Clinic, Danville, Pennsylvania; and 7Division of Nephrology and Hypertension, Mayo Clinic, Rochester, Minnesota

**ABSTRACT**

Background Mutations in PKD1 or PKD2 cause typical autosomal dominant polycystic kidney disease (ADPKD), the most common monogenic kidney disease. Dominantly inherited polycystic kidney and liver diseases on the ADPKD spectrum are also caused by mutations in at least six other genes required for protein biogenesis in the endoplasmic reticulum, the loss of which results in defective production of the PKD1 gene product, the membrane protein polycystin-1 (PC1).

Methods We used whole-exome sequencing in a cohort of 122 patients with genetically unresolved clinical diagnosis of ADPKD or polycystic liver disease to identify a candidate gene, ALG9, and in vitro cell-based assays of PC1 protein maturation to functionally validate it. For further validation, we identified carriers of ALG9 loss-of-function mutations and noncarrier matched controls in a large exome-sequenced population-based cohort and evaluated the occurrence of polycystic phenotypes in both groups.

Results Two patients in the clinically defined cohort had rare loss-of-function variants in ALG9, which encodes a protein required for addition of specific mannose molecules to the assembling N-glycan precursors in the endoplasmic reticulum lumen. In vitro assays showed that inactivation of Alg9 results in impaired maturation and defective glycosylation of PC1. Seven of the eight (88%) cases selected from the population-based cohort based on ALG9 mutation carrier state who had abdominal imaging after age 50; seven (88%) had at least four kidney cysts, compared with none in matched controls without ALG9 mutations.

Conclusions ALG9 is a novel disease gene in the genetically heterogeneous ADPKD spectrum. This study supports the utility of phenotype characterization in genetically-defined cohorts to validate novel disease genes, and provide much-needed genotype-phenotype correlations.

Received March 26, 2019. Accepted June 26, 2019.

Published online ahead of print. Publication date available at www.jasn.org.

Correspondence: Dr. Stefan Somlo, Section of Nephrology, Yale University School of Medicine, P.O. Box 208029, 333 Cedar Street, New Haven, CT 06520-8029, or Dr. Tooraj Mirshahi, Department of Molecular and Functional Genomics, Geisinger Clinic, 100 North Academy Avenue, Danville, PA 17822-2621. E-mail: stefan.somlo@yale.edu or tmirshahi@geisinger.edu

Copyright © 2019 by the American Society of Nephrology
detected (NMD) in either gene (ADPKD-NMD). Several genes which encode proteins in the endoplasmic reticulum (ER) that function in the biogenesis and quality control of integral membrane proteins have been implicated as disease genes for dominant polycystic kidney and liver phenotypes clinically characterized as isolated polycystic liver disease (PCLD) or ADPKD-NMD. PCLD, also known as autosomal dominant polycystic liver disease, presents with liver findings that are histologically and radiographically indistinguishable from those in ADPKD, but with few or absent kidney cysts. Although the ADPKD-PCLD diseases are inherited as dominant traits, accrual of somatic mutations resulting in biallelic loss of the respective disease genes in renal tubular and biliary epithelial cells is required for cyst initiation and accounts for the focal nature of the resulting polycystic phenotypes. Although there are a large number of client proteins that may be affected by loss of these ER-associated gene products, studies have shown that it is specifically the impaired post-translational maturation of the PKD1 gene product polycystin-1 (PC1) that produces(560,210),(688,229).

In this study, we identified ALG9 as a novel candidate gene in our unsolved cases for the PCLD–ADPKD-NMD phenotypic spectrum. ALG9 encodes α-1,2-mannosyltransferase, which catalyzes the addition of the seventh and ninth mannose molecules to a growing N-glycan precursor in the ER lumen. Autosomal recessive loss of ALG9 results in a severe congenital disorder of glycosylation (CDG) with a multigenic phenotype that includes kidney cysts. We validated ALG9 as a disease gene by in vitro functional bioassays showing both quantitative and qualitative defects in PC1 maturation after inactivation of ALG9. Amongst ALG9 loss-of-function mutation carriers identified in a large population-based cohort that had undergone whole exome sequencing (WES), we identified significant enrichment of kidney cysts.

**METHODS**

**PCLD–ADPKD-NMD Cohort and Exome Analysis**

This study is approved by the Institutional Review Board at Yale University. The studied cohort consists of 122 unrelated cases of PCLD or ADPKD-NMD enrolled at Yale or received de-identified from collaborators, as well as selected mild–moderate severity ADPKD-NMD cases from the Consortium of Radiologic Imaging Study of PKD and HALT clinical cohorts obtained from the National Institute of Diabetes and Digestive and Kidney Diseases Central Repository. Cases solved by pathogenic variants in established disease genes (PKD1, PKD2, PRKCSH, SEC6, GANAB, ALG8, SEC61B, PKHD1, and DNAJ-B1) have been excluded. Extraction of genomic DNA, WES, variant calling, and analysis for this cohort have been described previously. New samples were captured by the xGen Exome Research Panel reagent from Integrated DNA Technologies and sequenced on the Illumina HiSeq4000 platform. Variants were filtered to consider only rare heterozygous likely deleterious variants (stop gain, frameshift, canonical splice, missense with Combined Annotation Dependent Depletion [CADD] score >20) of high sequencing quality. Rare variants were considered to be those with minor allele frequency (MAF) of <0.001 in the gnomAD Browser. False positive variant calls as defined by PLDiff/depth <724 were excluded from consideration in all analyses. As no single gene was enriched with variants among cohort cases in statistical analyses at a genome-wide level, we ranked all genes represented by at least one loss-of-function variant based on the number of individuals carrying potentially deleterious variants in the gene. ALG9 was then selected from among top candidates by review of available literature and resources, including data on protein function and localization, and mouse or human phenotypes. ALG9 variants investigated as part of this study are deposited in ClinVar (SCV00091495, SCV00091494, SCV000930618-SCV000930621).

**In Vitro Gene Knockout**

The parental cell line noted as “control” in this study is a mouse renal tubule epithelial cell line which contains a three-copy Pkd1/7H bacterial artificial chromosome (BAC) with amino-terminal (N-terminal) FLAGx3 and carboxy-terminal (C-terminal) HAX3 epitope tags. The cell line has stable expression of Cas9 introduced by lentiviral transduction using lentCas9-Blunt (Addgene 52962). Guide-RNA sequence targeting the reverse strand of Alg9 exon 6 (5’-TCCGTCATGGCGAT-CAGCC-3’) was introduced transiently using electroporation of a modified pGL3-U6-sgRNA-PGK-puromycin plasmid (Addgene 51133) lacking the puromycin cassette to direct the Cas9 protein to make a double-stranded cut at this site. Transfected cells were clonally diluted after 72 hours and grown until sufficient quantity to passage. Cell line clones were screened by PCR and Sanger sequencing for frameshift mutations (Supplemental Figure 1).

**Significance Statement**

Dominantly inherited polycystic kidney and liver phenotypes occur when epithelial cells in these organs have reduction of polycystin-1 functional dosage. In a cohort of genetically unresolved polycystic liver and kidney disease, the authors identified heterozygous loss of function mutations in ALG9. ALG9 encodes an endoplasmic reticulum enzyme that builds N-glycans, and the authors show that Alg9 inactivation results in impaired polycystin-1 maturation. Using a novel ‘genotype-first’ approach to ascertain individuals based strictly on their ALG9 genotype from a large cohort of exome-sequenced individuals, the authors find that 7/8 (88%) of ALG9 mutation carriers over age 50 have multiple kidney cysts. Together, these findings identify ALG9 as a novel human polycystic kidney and liver disease gene and support the utility of a genotype-driven approach to candidate disease gene validation and genotype-phenotype correlation.
Immunoblotting and Glycosylation Analysis

Immunoblotting was performed on whole cell lysate prepared using RIPA buffer as reported previously. For quantitative comparisons, cell lysate concentration was measured by Lowry protein assay using standard reagents. For Ponceau staining, membrane was washed in Tris-buffered saline/Tween 20 (TBST), submerged in Ponceau S Solution (Sigma) for 10 minutes, rinsed briefly in water to remove background staining, and dried on tissue. Membrane was rehydrated in 100% methanol before proceeding with blocking in 5% nonfat milk in TBST. The following primary antibodies were used at 1:3000 dilution in 5% milk in TBST unless otherwise noted: anti-hemagglutinin (anti-HA) (3F10; Roche), anti–PC1-LRR (7e12, dilution 1:1000; Santa Cruz Biotechnology), chicken anti-green fluorescent protein (anti-GFP) antibody (GFP-1020; Aves Labs), and rabbit anti-Hsp90 antibody (Cell Signaling). For glycosylation studies, cell lysate obtained in RIPA buffer was treated with Endoglycosidase H (EndoH) per non-denaturing protocol (NEB) for 3 hours at 37°C.

Missense Variant Analysis

*In silico* analysis of missense variants was performed using prediction algorithms including CADD,^22^ MetaSVM and MetaLR,^25^ REVEL,^26^ and FATHMM-MKL.^27^ Biochemical validation studies were conducted on stable cell lines generated by lentiviral transduction of C-terminal GFP fusion with human ALG9 cDNA (ENST00000616540.4; 618 amino acid transcript) into the Alg9<sup>−/−</sup> cell line. Human ALG9 cDNA sequence was amplified from total RNA of human embryonic kidney 293 T cells using a two-step PCR strategy. The 5′ untranslated region to the middle of the transcript was amplified separately from overlapping region spanning from the middle of the transcript to the 3′ untranslated region, and these products were then used as a template for a second PCR reaction, using primers designed to add restriction sites to each end. This product was cloned into Zero Blunt TOPO vector using a cloning kit from ThermoFisher, and the correct clone was confirmed by Sanger sequencing. This ALG9 sequence in the TOPO vector was modified using a site-directed mutagenesis protocol. Briefly, Pfu polymerase (Promega) amplified a modified plasmid primed by complimentary forward and reverse primers containing the sequence edit (Supplemental Methods). The resultant product was transformed into DH5<sup>α</sup> competent cells after digestion of parental template plasmid with DpnI (NEB). A single modified clone for each missense variant was confirmed by sequencing, cut by restriction digest, ligated into pLVX-IRE-Puro (Clontech), and again each was confirmed with Sanger sequencing. GFP expression and localization for each cell line was examined by immunoblotting and fluorescence microscopy, respectively.

MyCode Cohort Exome Analysis

Whole-exome sequencing and variant calling were performed before this study on 92,455 MyCode Community Health Initiative participants as part of the Geisinger-Regeneron DiscovEHR collaboration. Methods for whole-exome capture and variant calling in this cohort have been previously described. Briefly, NimbleGen (SeqCap VCRome) probes were used for exome capture of the first approximately 61,000 samples and a modified version of Integrated DNA Technologies xGen probes was used for the remaining approximately 31,000 samples. Variants were annotated using the Ensembl Variant Effect Predictor with Ensembl 90 definitions. PLDiff/depth and gnomAD were used as described above to filter out common and false positive variants. Carriers of ALG9 loss-of-function variants were identified from this cohort and confirmed with Sanger sequencing. We accessed pedigree reconstruction data assembled from exome-based identity-by-descent analysis to determine the relatedness between individuals to identify any first or second degree relatives for the ALG9 carriers in the cohort.

MyCode Cohort Case versus Control Study

The study of the MyCode cohort was approved by the Institutional Review Board at Geisinger. Patients carrying the ALG9 mutation who were eligible for matching were defined as those with abdominal imaging studies available for digital review with a definable native kidney eGFR—i.e., not on dialysis or with kidney transplant. One imaging exam was selected for each case by a predefined algorithm: magnetic resonance imaging (MRI) > computed tomography (CT) > ultrasound, and exams with contrast > those without. Clinical information for carriers of ALG9 mutations was obtained from clinical databases and used to select random matched controls in a 2:1 ratio (see power calculation below) based on six parameters (see Results). A radiologist with expertise in abdominal imaging (W.J.T.), blinded to patient genotypes and clinical information, provided interpretation of kidney and liver findings in the selected imaging studies using clinical radiologic diagnostic criteria. Peripelvic cysts, although infrequent (data not shown), were excluded from the analysis given their lymphatic origin and inability to reliably delineate discrete cysts from the collecting system. For CT scans, a cyst was defined as a homogeneous mass of water attenuation (−10 to 20 HU) with an imperceptible wall and no enhancement after administration of contrast or, in the case of hemorrhagic or proteinaceous cysts, as a homogeneus, nonenhancing lesion with precontrast density >20 HU. Size thresholds were set based on the heterogeneity of the CT data, discussed below (see Results). For MRI, a cyst was defined as a lesion distinct from the surrounding parenchyma with either homogenous internal T2-weighted signal intensity consistent with simple fluid, or a lesion with T1 and T2 signal characteristics consistent with a hemorrhagic or proteinaceous cyst. In cases where postcontrast imaging was available, the lesions showed no enhancement. For ultrasound imaging, a cyst was defined as an anechoic structure with a distinct posterior wall and posterior enhancement.

Statistical Analyses

A two-tailed test was used to compare means of biologic repeats on Western blots. Sample size calculations for case
versus control analysis were performed using Power Analysis Sample Size (PASS) with the goal of achieving 70% power in detecting a difference if cases and controls had an 80% and very conservatively selected 20% incidence of kidney cysts, respectively. Case versus control comparisons were performed using the Fisher exact test.

**RESULTS**

**ALG9 Is a Candidate Gene for PCLD–ADPKD-NMD**

We performed WES on 122 unrelated patients with ADPKD-NMD or PCLD without pathogenic mutations in previously established nor proposed disease genes. We identified one case with a heterozygous loss-of-function mutation in exon 6 of **ALG9**, p.W227X, and five other cases with rare (MAF<1×10^{-3}) heterozygous missense variants of unknown consequence in **ALG9** (Table 1). Each of the five missense variants was predicted to be in the top 0.3% most likely deleterious of all potential genome single nucleotide variants (CADD-phred ≥25) but had differing predictions with other in silico algorithms (Table 1). 33

In vitro inactivation of the previously reported PCLD–ADPKD-NMD disease genes that encode ER proteins have shown significant quantitative reduction and qualitative alteration of mature PC1 protein. 6,8,13 In vitro biallelic inactivation models the cellular recessive genotype resulting from somatic second hits in cyst-forming epithelium. We therefore investigated the effects of inactivation of **ALG9** on PC1 in a cell culture system. The full-length PC1 protein (PC1-FL) is normally cleaved into an N-terminal fragment and an 11-transmembrane-spanning C-terminal fragment (PC1-CTF) before leaving the ER and trafficking to functional destinations on the cell surface. We used CRISPR/Cas9 to introduce a frameshift in exon 6 of **ALG9** to inactivate it in a mouse kidney tubular epithelial cell line. The isogenic unmodified parental cell expressing Cas9 was used as control. The parental cell line expresses three copies of a BAC transgene, **Pkd1**/-/-BAC, containing the full genomic sequence of mouse **Pkd1** modified with addition of N-terminal FLAGx3 and C-terminal HAx3 epitope tags. 7,15 Sanger sequencing of the genomic region of **ALG9** targeted by the guide RNA in a clonally selected candidate **ALG9**−/− cell line confirmed the presence of a homozygous frameshifting mutation (Supplemental Figure 1). **ALG9**−/− cells had no apparent morphologic changes on light microscopy compared with the parental cell line. **ALG9**−/− cells showed a significant decrease in the steady-state level of PC1-CTF to approximately 70% of the control intensity, quantified in four biologic samples per genotype by anti-HA (Figure 1A, left panel; Supplemental Figure 2). A total of 57 of the 60 predicted N-glycosylation sites on PC1 are in the N-terminal fragment. After exit from the ER, enzymes in the Golgi modify the N-glycans conferring resistance to EndoH which is associated with slower migration on protein gel electrophoresis. This allows experimental distinction of the immature EndoH-sensitive PC1 N-terminal fragment (PC1-NTS) that has not traversed the Golgi from the EndoH-resistant fragment (PC1-NTR) that reaches the cell surface. Immunoblot analysis with antibody directed to the PC1 N-terminal resolves three fragments: PC1-FL, PC1-NTR, and PC1-NTS (Figure 1A, right panel). **ALG9**−/− cells show quantitatively decreased PC1-NTR, indicating reduced mature, cell surface–expressed PC1. In addition to quantitative differences, PC1-FL and PC1-NTS from **ALG9**−/− cells have faster migration in denaturing gel electrophoresis compared with control cells (Figure 1A, Supplemental Figure 2). This may also be true of PC1-NTR, but the

---

**Table 1. ALG9 heterozygous mutation carriers in 122 genetically unresolved patients with PCLD–ADPKD-NMD**

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Nucleotide Change</th>
<th>Amino Acid Change</th>
<th>MAF gnomAD</th>
<th>CADD</th>
<th>MetaSVM</th>
<th>MetaLR</th>
<th>REVEL</th>
<th>FATHMM-MKL</th>
<th>Age/ Gender</th>
<th>Kidney Cysts</th>
<th>Liver Cysts</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2618139</td>
<td>c.681G&gt;A</td>
<td>p.W227X</td>
<td>Novel</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.912</td>
<td>D</td>
<td>60/M</td>
<td>Innumerable</td>
<td>N/A</td>
</tr>
<tr>
<td>YU202</td>
<td>c.1109G&gt;A</td>
<td>p.R370K</td>
<td>4×10^{-6}</td>
<td>33</td>
<td>D</td>
<td>D</td>
<td>0.912</td>
<td>D</td>
<td>56/M</td>
<td>7:11</td>
<td>Innumerable</td>
</tr>
<tr>
<td>YU229</td>
<td>c.694G&gt;C</td>
<td>p.A232P</td>
<td>3×10^{-4}</td>
<td>25</td>
<td>T</td>
<td>T</td>
<td>0.395</td>
<td>D</td>
<td>67/F</td>
<td>N/A</td>
<td>Multiple</td>
</tr>
<tr>
<td>YU36b</td>
<td>c.839C&gt;T</td>
<td>p.A280V</td>
<td>7×10^{-6}</td>
<td>31</td>
<td>T</td>
<td>T</td>
<td>0.324</td>
<td>D</td>
<td>66/F</td>
<td>—</td>
<td>Multiple</td>
</tr>
<tr>
<td>453976c</td>
<td>c.944A&gt;G</td>
<td>p.N315S</td>
<td>6×10^{-5}</td>
<td>25</td>
<td>T</td>
<td>T</td>
<td>0.574</td>
<td>D</td>
<td>42/F</td>
<td>Innumerable</td>
<td>Innumerable</td>
</tr>
<tr>
<td>YU394</td>
<td>c.1550G&gt;T</td>
<td>p.R517L</td>
<td>4×10^{-6}</td>
<td>27</td>
<td>T</td>
<td>T</td>
<td>0.462</td>
<td>D</td>
<td>63/F</td>
<td>1:0</td>
<td>Multiple</td>
</tr>
</tbody>
</table>

Yale case cohort comprised of PCLD and mild-moderate ADPKD-NMD cases from Consortium of Radiologic Imaging Study of PKD (CRISP)/HALT Progression of Polycystic Kidney Disease (HALT PKD) cohorts.

N/A, not applicable; D, deleterious; T, tolerated.

aENST00000616540.4 (CCDS73380).

bVariant prediction (see Methods).

cREVEL scores range from zero to one, with one representing all simulations suggest pathogenicity.

dCyst counts (right-left) or description from available imaging.

3HALT identifier.

4Mayo Clinic Family T57.

5Mayo Clinic Family T39.

6Mayo Clinic Family T17.

7CRISP identifier.
overlapping migration of PC1-FL and PC1-NTR from Alg9<sup>−/−</sup> cells prevents definitive conclusion. The genotype-dependent differences in migration of both cleaved and uncleaved PC1 suggest that, in the absence of Alg9, there is altered glycosylation of PC1. Enzymatic removal of N-glycans, which on PC1-FL are all EndoH sensitive, results in identical migration between genotypes (Figure 1B), confirming that the faster migration of PC1-FL from Alg9<sup>−/−</sup> cells is due to altered N-glycosylation. Finally, both the altered protein expression levels and altered glycosylation of PC1 in Alg9<sup>−/−</sup> cells showed expression indistinguishable from control cells when human ALG9-GFP is re-expressed.
in Alg9<sup>−/−</sup> cells (Figure 1A). This confirms that the effects on PC1 maturation are a specific consequence of loss of Alg9 and provides strong support for the conclusion that ALG9 is a human polycystic disease gene.

**Assessment of ALG9 Missense Variants Found in Patients with PCLD–ADPKD-NMD**

We next used this system to evaluate the functional consequences of the ALG9 missense variants identified in our patients (Table 1). We expected that re-expression of benign human ALG9 missense variants would rescue PC1 properties in patients (Table 1). We expected that re-expression of benign ALG9 missense variants would rescue PC1 properties in ALG9<sup>−/−</sup> cells, whereas deleterious missense variants would fail to do so. We modified the human wild-type ALG9-GFP cDNA sequence to introduce each of the five missense mutations (Table 1). In addition, we produced a clone with a known deleterious missense variant ALG9-p.Y287C<sup>13</sup> as a positive control. Each cDNA was introduced into the Alg9<sup>−/−</sup> cell line by lentiviral transduction and stably expressing cell lines were selected. Expression of each clone was verified by GFP epifluorescence in cell culture (data not shown). Alg9<sup>−/−</sup> cells expressing the positive control ALG9-p.Y287C, as well as the line expressing the patient-derived variant ALG9-p.R370K, showed quantitatively reduced PC1-CTF and slower migrating PC1-FL indistinguishable from that seen in the Alg9<sup>−/−</sup> (Figure 1C). The Alg9<sup>−/−</sup> cells expressing ALG9-p.A232P, ALG9-p.A280V, ALG9-p.N315S, or ALG9-p.R517L each showed complete rescue of PC1 quantitative and qualitative maturation defects (Figure 1D). The data indicate that, among the rare missense mutations found in the patients, only ALG9-p.R370K resulted in ALG9 loss of function (Table 1). The demonstration of significant impairment of PC1 maturation resulting from ALG9 exon 6 truncation or ALG9-p.R370K, and the occurrence of these variants in patients with the ADPKD-PCLD phenotype support these as the causative variants for these two patients in our cohort (Table 1). YU202, with pathogenic missense variant ALG9-p.R370K, is a 56-year-old male with clinically diagnosed PCLD (Figure 1E). His father also had PCLD, but no genetic diagnosis. G8261813, with ALG9-p.W227X, was diagnosed with ADPKD-NMD as an enrolled participant in the HALT PKD study (imaging not available). These clinical data indicate that ALG9-related disease can span the spectrum from PCLD to ADPKD.

**Genotype-First Approach To Characterize ALG9 Mutation Carriers**

To evaluate these findings further, we determined the burden of polycystic kidney and liver phenotypes in cases selected based solely on the presence of ALG9 loss-of-function mutations from a large population-based cohort in which WES data are linked to electronic health record data.<sup>30</sup> Among the 92,455 exomes queried, we identified 21 carriers of rare (MAF<1×10<sup>−3</sup>) heterozygous ALG9 loss-of-function mutations (nonsense, frameshift, canonical splice, or start loss) meeting sequence quality control thresholds (see Methods). There were no instances of individuals with homozygous or compound heterozygous ALG9 loss-of-function mutations. Principle component analysis and relatedness analysis of exome data indicated that all 21 of these patients were of European descent and no close familial relations were present between them. Of the 21 patients, 14 (67%) had at least one abdominal imaging report available for evaluation. The ALG9 variants (Table 2) in these 14 individuals were confirmed by Sanger sequencing. None carried loss-of-function mutations in established cystic disease genes, and missense variants in these genes were benign by in silico prediction (Supplemental Table 1). Clinical radiology reports described kidney cysts (n=7) or hypodensities suggestive of cysts (n=4) in these 14 patients, and liver cysts in only one. To characterize and quantify these findings and determine the specificity of these phenotypes to the ALG9 genotype in light of the sporadic occurrence of kidney and liver cysts in some individuals with increasing age, we designed a case versus control analysis in which a radiologist blinded to case-control status evaluated and quantitated cyst burden.

**Case versus Matched Control Analysis**

We included 11 of the 14 ALG9 cases for the case-control study. The excluded patients include two for which only a scanned radiology report but no digital imaging was available for re-evaluation (MC6, MC13; Table 2). The third patient had reached ESKD before first available imaging (MC11; Table 2). We selected two random controls for each of the 11 remaining cases from exome-sequenced individuals lacking any coding region variants in PKD1 or PKD2 (MAF<2×10<sup>−5</sup>) or other established PCLD gene (MAF<1×10<sup>−2</sup>); they were matched for six parameters: age at time of imaging, type of imaging study including use of contrast, year of imaging, CKD stage at time of imaging, gender, and ethnicity. We defined four or more cysts in patients aged 50 years or more as a threshold for positive evidence of cystic disease. This criterion was informed by Rule et al.<sup>34</sup> who report one, two, or three cysts of ≥5 mm seen in 26%, 9.8%, and 4.3%, respectively, in healthy organ donors aged 50 or more. Four or more cysts only occurred 1.2% of those healthy individuals.<sup>34</sup> We assessed kidney and liver lesions in a blinded fashion using prespecified radiographic criteria (see Methods) for all 33 individuals (11 cases, 22 matched controls; Supplemental Table 2; Table 2). Although matched between cases and controls, the 33 imaging studies included different imaging modalities with variable technical parameters inherent in a cohort spanning multiple decades. In aggregate, the CT exams, which were the majority of the data, dictated that a size threshold of ≥8 mm allowed for a definitive application of the specified criteria for a cyst, and a minimum size threshold of ≥4 mm allowed for lesions to be detectable across all exams. As such, lesions ≥8 mm were classified as either “cyst” or “indeterminate mass,” and lesions <8 mm but ≥4 mm were reported as “too small to characterize” (TSTC) per common nomenclature. To facilitate direct comparison to the CT data, the same TSTC thresholds were applied to the MRI data. In contrast to CT and MRI, the concept of TSTC does not
Table 2. Alg9 loss-of-function mutations carriers with abdominal imaging among 92,455 patients in a health system

<table>
<thead>
<tr>
<th>Identifier/Gender</th>
<th>Nucleotide Change(^a)</th>
<th>Amino Acid Change</th>
<th>MAF gnomAD</th>
<th>Imaging Type (age)(^b)</th>
<th>Kidney Cysts(^c)</th>
<th>TSTC(^c)</th>
<th>Nephrolithiasis(^d)</th>
<th>eGFR (age)(^e)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kidney cysts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC2/F</td>
<td>c.427C&gt;T</td>
<td>p.R143X</td>
<td>Novel MRI</td>
<td>(65)</td>
<td>10</td>
<td>18</td>
<td>Y</td>
<td>60 (65)</td>
</tr>
<tr>
<td>MC5/F</td>
<td>c.1018+1G&gt;A</td>
<td>Splice variant</td>
<td>Novel CT+</td>
<td>(86)</td>
<td>39(^f)</td>
<td>51</td>
<td>Y(^g)</td>
<td>38 (96)</td>
</tr>
<tr>
<td>MC7/M</td>
<td>c.1472delA</td>
<td>p.N491fsTer33</td>
<td>Novel CT+</td>
<td>(69)</td>
<td>6 (9)</td>
<td>22</td>
<td>Y</td>
<td>74 (73)</td>
</tr>
<tr>
<td>MC10/M</td>
<td>c.883_885delICCT_insGTAAA</td>
<td>p.P295VsTer13</td>
<td>Novel CT+</td>
<td>(68)</td>
<td>9</td>
<td>21</td>
<td>Y</td>
<td>52 (75)</td>
</tr>
<tr>
<td>MC12/F</td>
<td>c.511C&gt;T</td>
<td>p.R171X</td>
<td>8.13E−06</td>
<td>CT+ (61)</td>
<td>4 (9)</td>
<td>20</td>
<td>—</td>
<td>46 (63)</td>
</tr>
<tr>
<td><strong>One or fewer cysts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC1/M</td>
<td>c.1018+1G&gt;A</td>
<td>Splice variant</td>
<td>Novel CT+</td>
<td>(42)</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td>50 (51)</td>
</tr>
<tr>
<td>MC4/F</td>
<td>c.1088_1091delTCCA</td>
<td>p.I363FsTer52</td>
<td>4.06E−06</td>
<td>CT+ (82)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>47 (85)</td>
</tr>
<tr>
<td>MC8/F</td>
<td>c.3G&gt;A</td>
<td>Start loss</td>
<td>Novel CT+</td>
<td>(37)</td>
<td>—</td>
<td>1</td>
<td>—</td>
<td>72 (44)</td>
</tr>
<tr>
<td>MC14/F</td>
<td>c.3G&gt;A</td>
<td>Start loss</td>
<td>Novel MRI+</td>
<td>(37)</td>
<td>1</td>
<td>3</td>
<td>+</td>
<td>90 (44)</td>
</tr>
<tr>
<td><strong>Cases not matched for case-control analysis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC6/M</td>
<td>c.1018+1G&gt;A</td>
<td>Splice variant</td>
<td>Novel CT+</td>
<td>(76)</td>
<td>Y(^h)</td>
<td>—</td>
<td>—</td>
<td>47 (78)</td>
</tr>
<tr>
<td>MC11/F</td>
<td>c.1363C&gt;T</td>
<td>p.R455X</td>
<td>4.06E−06</td>
<td>CT (70)</td>
<td>Y(^h),(^i)</td>
<td>—</td>
<td>ESKD (66)</td>
<td>—</td>
</tr>
<tr>
<td>MC13/F</td>
<td>c.566-1G&gt;A</td>
<td>Splice variant</td>
<td>1.22E−05</td>
<td>CT+ (49)</td>
<td>Y(^h)</td>
<td>—</td>
<td>—</td>
<td>97 (50)</td>
</tr>
</tbody>
</table>

US, ultrasound.

\(^a\) ENST00000616540.4 (CCDS73380).

\(^b\)+, with contrast.

\(^c\) Kidney cysts (>8 mm) and lesions TSTC (4–8 mm) as described in Methods and Results. When additional imaging allowed for recharacterization of indeterminate masses as cysts, cyst count inclusive of these is noted in parentheses after count from originally selected image.

\(^d\) Noted during blinded analysis. Y, yes; +, nephrolithiasis noted on additional CT scan if available.

\(^e\) Most recent outpatient eGFR.

\(^f\) Among cases with liver imaging, this was the only case with liver cysts noted.

\(^g\) Pathology of bilateral nephrolithiasis demonstrated uric acid composition.

\(^h\) Based on clinical radiology report.

\(^i\) Native kidneys by CT 2/4 yr after transplant/ESKD.
apply to the ultrasound data, as such findings would not be reliably detected by that modality.

Among the 11 ALG9 mutation carriers (cases) analyzed in the case-control analysis, seven (64%) had four or more kidney cysts (Figure 2, Table 2). None of the 22 controls met this threshold ($P<0.001$; Tables 2 and 3; see Methods). The four ALG9 cases with less than four kidney cysts included the three youngest patients, aged 37, 37, and 42, each of whom also had at least one TSTC in addition to the observed cysts. The fourth case was an 82-year-old female with no cysts. Considering only the cases and controls $\geq 50$ years of age, seven of eight cases (88%) versus none of 16 controls had four or more kidney cysts ($P<0.001$; Table 3). The background occurrence of a single cyst or TSTC was relatively common in controls (ten of 22; 45%), but significantly less than in cases (ten of 11; 91%; $P=0.02$). The case versus control cyst count comparison for the selected threshold of four or more kidney cysts remained robustly significant ($P<0.01$) even when criteria were modified to include TSTCs or any number of bilateral cysts (Table 3).

The severity of the cystic phenotype in cases is highly variable between individuals. Cyst numbers ranged from zero to innumerable kidney cysts (Figure 2, Table 2). In contrast, no controls had more than two cysts. Cysts were bilateral in all seven affected cases, and were mentioned in the clinical

![Figure 2](image-url). Cystic kidney phenotypes in carriers of ALG9 heterozygous loss-of-function variants are of varying severity. (A) Ultrasound long-axis views from anteroposterior sweep, and (B and C) axial CT scan sections. Quantification of lesions in all cases is presented in Table 2.

### Table 3. Case versus matched control analysis

<table>
<thead>
<tr>
<th>Cyst count</th>
<th>Cases</th>
<th>Controls</th>
<th>$P$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case-control cohort</td>
<td>n=11</td>
<td>n=22</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cyst count $\geq 4$</td>
<td>7</td>
<td>0</td>
<td>0.0055</td>
</tr>
<tr>
<td>Cysts plus TSTC combined count $\geq 4$</td>
<td>8</td>
<td>4</td>
<td>0.0021</td>
</tr>
<tr>
<td>Bilateral cysts*</td>
<td>7</td>
<td>2</td>
<td>7.5×10^{-4}</td>
</tr>
<tr>
<td>Subset of age over 50</td>
<td>n=8</td>
<td>n=16</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cyst count $\geq 4$</td>
<td>7</td>
<td>0</td>
<td>0.0087</td>
</tr>
<tr>
<td>Cysts plus TSTC combined count $\geq 4$</td>
<td>7</td>
<td>4</td>
<td>7.5×10^{-4}</td>
</tr>
<tr>
<td>Bilateral cysts*</td>
<td>7</td>
<td>2</td>
<td>7.5×10^{-4}</td>
</tr>
</tbody>
</table>

* Bilateral cysts: at least one cyst in each kidney.
radiology reports for all seven cases. Five of the six cases with a finite number of cysts had 18 or more lesions of 4–8 mm diameter, suspected to be early cysts but noted as TSTCs (Figure 2, Table 2). The largest kidney cysts in cases ranged from 1.2 to 4.8 cm. For two cases with four and six cysts respectively, specific additional kidney lesions counted as indeterminate masses could be categorically defined to be cysts upon review of additional imaging studies (Table 2). The liver was not imaged in the two ALG9 carriers who only had renal ultrasound available; only one of the nine cases with liver imaging (MC5) had liver cysts. MC5’s only partially imaged liver contained two cysts, up to 1.1 cm in size, and four TSTCs.

There was limited availability of familial data in this study. At the time of YU202’s enrollment, he reported that his father had multiple liver cysts at age 83, and mother was living and well. For the cases in the MyCode cohort, we used the available pedigree reconstruction from the 92,455 individuals to supplement the absence of relevant family history data in clinical charts. Only MC1 and MC3 have relatives in the cohort, a mother and a brother, respectively, and each are noncarriers of the respective ALG9 mutations. MC3 (Figure 2A, Table 2) is the most significantly affected case in this cohort, and notably the brother of this individual did not share the ALG9 mutation and did not have kidney cysts on a CT scan performed at age 70.

The blinded review incidentally noted nephrolithiasis in four ALG9 mutation carrier cases (two CT scan, one ultrasound, and one MRI) on the single imaging study used in the study. This was only noted in one control (ultrasound exam). As MRI and ultrasound are inferior for identifying nephrolithiasis, and nephrolithiasis can be transient on imaging, we performed additional chart review to consider other available CT scans for the cases and controls. CT scan clinical reports confirmed the nephrolithiasis found in the four cases, and reported nephrolithiasis in an additional three cases for a total of seven of 11 (64%) cases with digital imaging (Table 2). Six of these seven were among the cases with multiple kidney cysts and the seventh had one kidney cyst and three TSTCs. The most recent CT scan clinical reports for the controls did not describe any additional cases of nephrolithiasis, however, we did not review every available scan for each control as this study was not designed to make a case versus control comparison for this finding. Of the remaining ten ALG9 mutation carriers who lacked digital imaging data, there was no notation of nephrolithiasis by International Classification of Diseases code, an insensitive measure, nor on the imaging report for MC6, MC11, or M13.

In aggregate, this genotype-first analysis shows a significantly greater incidence of multiple kidney cysts in ALG9 loss-of-function carriers compared with individuals without this genotype. Given the expanding availability of exome-sequenced cohorts tied to the live electronic health record, this genotype-first approach will be a valuable adjunct in validating the pathogenicity of candidate genes for rare Mendelian disease traits.

**DISCUSSION**

Our study is the first to complement gene discovery in clinically ascertained polycystic kidney and liver disease with phenotypic characterization in genetically defined at-risk patients from population-based exome sequencing efforts. We show the significant benefits of combining these complementary strategies. Recent implication and analyses of disease genes for ADPKD and PCLD demonstrate that these clinically defined human phenotypes are found in individuals with heterozygous mutations in genes required for sufficient PC1 functional dosage. As such, the implication and study of novel disease genes for this genetically heterogeneous phenotype offers concrete entry points for novel biologic investigations into PC1 function, an entity of central importance to the design of targeted therapies for ADPKD and PCLD–ADPKD–NMD. These phenotypes exist both mechanistically and phenotypically on more of a continuum than was realized before these studies.6–8,13 This study adds the novel benefit of being able to detail findings in mutation carriers of a disease gene independent from the influence of the clinical polycystic cohort from which it was discovered.

The population-based ascertainment of ALG9 mutation carriers in the MyCode cohort provides several insights. Firstly, because human variation assures that no additional gene variants would be over-represented in the carriers of ten novel or ultra-rare ALG9 variants, it provides the strength of a randomized study investigating the effect of the ALG9 variant alone, not requiring a co-occurring variant. We found that the penetrance of four or more definitive kidney cysts in individuals over age 50 with ALG9 mutations is 88% (seven of eight). The size and number of cysts found in ALG9 carriers is generally milder than typical ADPKD, showing large variation in expressivity. ALG9 carriers show a kidney-predominant phenotype, differing from hypothesis based on four of the six PCLD–ADPKD–NMD disease genes encoding proteins in the ER which have been described in patients exclusively with PCLD. This raises the possibility that similar genotype-first analyses for other PCLD genes related to ER biogenesis may establish whether the stronger correlation with PCLD is a reflection of the disease mechanism or related to the criteria for ascertainment used in previous discovery cohorts.3–5,7 The generally mild phenotype in human ALG9 disease is likely determined by two factors: the relative frequency of somatic second hit mutations inactivating the normal copy of ALG9, and the level of damage to the PC1 functional dose that results from loss of ALG9.

ALG9 joins two other genes, ALG8 and PMM2, implicated in kidney and liver cyst phenotypes whose recessive loss results in rare CDG.7,35 Kidney cystic disease attributed to PMM2 mutations is described as part of a hyperinsulinemic hypoglycemia and polycystic kidney disease syndrome attributed to recessively inherited biallelic nontruncating mutations, at least one of which is a unique promoter variant proposed to result in decreased expression of PMM2 specifically in the pancreas, kidney, and liver, avoiding the typical neurologic symptoms of PMM2-CDG.35 Due to a germline recessive genotype, the
kidney cystic phenotype in hyperinsulinemic hypoglycemia and polycystic kidney disease does not require accrual of somatic second hit mutations, is diagnosed at a younger age, and affects the organ diffusely. Recessive loss of the α-1,2-mannosyltransferase encoded by ALG9 causes the rare ALG9-CDG. Unique from other CDGs, kidney cysts are present in all five of five cases in which kidney examination was included among the 12 reported cases of ALG9-CDG. The pathogenic ALG9 missense variant p.R370K described in this report, as well as the variants that cause ALG9-CDG (p.Y287C, p.N527K, p.E530K; ENST00000616540.4; CCDS73380), are all predicted to be on the ER luminal side of the ALG9 ER transmembrane protein.16,19,36 ALG9 functions to add specific mannose molecules to the assembling N-glycan precursors in the ER lumen.17 The presence of kidney cysts in ALG9-CDG at birth, in contrast to the age-dependent and incomplete penetrance in patients heterozygous for ALG9, is supportive of the hypothesis that a cellular recessive mechanism established for ADPKD and PCLD is required to initiate cysts due to ALG9 mutations. As the proteins encoded by ALG9, ALG8, and PMM2 are enzymes with nonredundant roles in the assembly of N-glycans, our findings suggest a requirement of N-glycosylation in renal tubular and biliary epithelial homeostasis.

ALG9 is now the third ER protein—encoding polycystic disease gene, after GANAB and DNAJB11, described to explain either kidney-predominant polycystic phenotypes or PCLD.6,8 We also reported a young female with a mutation in ALG8 who had eight bilateral kidney cysts without liver cysts.7 DNAJB11 is the only of these genes with reported cases of ESKD, but this was proposed to result from tubulointerstitial fibrosis rather than polycystic kidney disease.8 In this study, one ALG9 mutation carrier reached ESKD at age 66 with atrophic native kidneys containing bilateral renal cysts on imaging 4 years after onset of ESKD. Lack of records from earlier time points precludes drawing any conclusions in this case. In the two ALG9 cases with the largest number of kidney cysts (MC3, MC5), the eGFRs were 26 and 38, and the kidney size was mildly enlarged or normal at 14/15 and 10/12 cm, respectively. The patient we report from the HALT ADPKD cohort, for whom imaging and organ size measurements were not available, had an eGFR of approximately 30 ml/min at age 60. It remains possible that ALG9 falls in the group of ER-associated polycystic kidney disease genes that may manifest this way. As is the case for PCLD, it also allows for quantification of additional imaging findings such as nephrolithiasis. This approach required the initial implication of ALG9 in phenotype-based cohorts. Although our data describe polycystic kidneys and suggest that a liver-specific (PCLD) phenotype may be considered when mutations in PKD1 or PKD2 are not found in mild–moderate cases of ADPKD-NMD, as well as for PCLD and phenotypes spanning these clinical distinctions. We suggest the genotype-based ascertainment approach be applied to other diseases with extensive genetic heterogeneity to help validate candidate genes and better inform genotype-phenotype correlations and prognostic advice in the era of precision medicine.

ACKNOWLEDGMENTS

We thank Dr. F. Perry Wilson and Dr. Brandon K. Fornwalt for advice on study design.

Dr. Besse performed experiments and cowrote the manuscript.

Dr. Chang, Mr. Luo, Dr. Triffo, Ms. Moore, Dr. Gulati, and Mr. Hartzel analyzed data. Dr. Mane provided exome sequencing. Regeneron Genetics Center (RGC) provided access to exome sequencing data. Dr. Torres contributed patient samples and data. Dr. Besse, Dr. Chang, Mr. Luo, Dr. Triffo, Dr. Somlo, and Dr. Mirshahi designed the study. Dr. Somlo and Dr. Mirshahi cowrote the manuscript.

The list of contributors from RGC and Yale Center for Mendelian Genomics (CMG) is provided in the Supplemental Acknowledgments.

DISCLOSURES

Dr. Torres reports grants and honoraria from Otsuka Pharmaceuticals, grants from Palladio Biosciences, Mironid, Sanofi Genzyme, Acceleron Pharma Inc., Regulus Therapeutics, and Blueprint Medicines, all outside the submitted work. Dr. Somlo reports personal fees from Goldfinch Bio, outside the
subMITTED WORK. Dr. Mirshahi reports grants from National Institutes of Health (NIH) National Institute of General Medical Sciences and grants from NIH National Human Genome Research Institute, during the conduct of the study.

FUNDING

This work was supported by PKD Foundation Research Grant and Fellowships to Dr. Besse (217G18a and 190F15a) and Dr. Gulati (207F17a); NIH grants DK100592 and DK051041 to Dr. Somlo, DK090728 to Dr. Torres, GM111913 to Dr. Mirshahi, and DK106513-01 to Dr. Chang; the George M. O’Brient Kidney Center at Yale (P30 DK079310); the Mayo Clinic Robert M. and Billie Kelley Pirnie Translational PKD Center; and the Yale CGM (NIH M#UM1HG006504-05).

SUPPLEMENTAL MATERIAL

This article contains the following supplemental material online at http://jasn.asnjournals.org/lookup/suppl/doi:10.1681/ASN.2019030298/-/DCSupplemental.

Supplemental Table 1. Additional coding region genetic variants with MAF <1% by whole exome sequencing in published ADPKD-PCLD genes in ALG9 loss-of-function carriers.

Supplemental Table 2. Matched Control Phenotype Data.

Supplemental Figure 1. Sanger sequencing of Alq9<sup>–/–</sup> cell line.

Supplemental Figure 2. Qualification of relative PC1 protein expression level in Alq9<sup>–/–</sup> cells.

Supplemental Methods.

Supplemental Acknowledgments.

REFERENCES
