NIH Public Access
Author Manuscript
Mol Genet Metab. Author manuscript; available in PMC 2011 February 1.

Published in final edited form as:
Mol Genet Metab. 2010 February ; 99(2): 160. doi:10.1016/j.ymgme.2009.10.010.

PKHD1 Sequence Variations in 78 Children and Adults with Autosomal Recessive Polycystic Kidney Disease and Congenital Hepatic Fibrosis

Meral Gunay-Aygun[1], **Maya Tuchman**[1], **Esperanza Font-Montgomery**[1], **Linda Lukose**[1], **Hailey Edwards**[1], **Angelica Garcia**[1], **Surasawadee Ausavarat**[1], **Shira G. Ziegler**[1], **Katie Piwnica-Worms**[1], **Joy Bryant**[1], **Isa Bernardini**[1], **Roxanne Fischer**[1], **Marjan Huizing**[1], **Lisa Guay-Woodford**[4], and **William A. Gahl**[1]

[1] Medical Genetics Branch, National Human Genome Research Institute, Bethesda, MD

[2] National Institutes of Health Clinical Center, Bethesda, MD

[3] Molecular Imaging Program, National Cancer Institute, Bethesda, MD

[4] University of Alabama, Birmingham, AL

Abstract

PKHD1, the gene mutated in autosomal recessive polycystic kidney disease (ARPKD)/Congenital hepatic fibrosis (CHF), is an exceptionally large and complicated gene that consists of 86 exons and has a number of alternatively spliced transcripts. Its longest open reading frame contains 67 exons that encode a 4074 amino acid protein called fibrocystin or polyductin. The phenotypes caused by *PKHD1* mutations are similarly complicated, ranging from perinatally-fatal PKD to CHF presenting in adulthood with mild kidney disease. To date, more than 300 mutations have been described throughout *PKHD1*. Most reported cohorts include a large proportion of perinatal-onset ARPKD patients; mutation detection rates vary between 42% and 87%. Here we report *PKHD1* sequencing results on 78 ARPKD/CHF patients from 68 families. Differing from previous investigations, our study required survival beyond 6 months and included many adults with a CHF-predominant phenotype. We identified 77 *PKHD1* variants (41 novel) including 19 truncating, 55 missense, 2 splice, and 1 small in-frame deletion. Using computer-based prediction tools (GVGD, PolyPhen, SNAP), we achieved a mutation detection rate of 79%, ranging from 63% in the CHF-predominant group to 82% in the remaining families. Prediction of the pathogenicity of missense variants will remain challenging until a functional assay is available. In the meantime, use of *PKHD1* sequencing data for clinical decisions requires caution, especially when only novel or rare missense variants are identified.

Keywords

PKHD1; autosomal recessive polycystic kidney disease; congenital hepatic fibrosis; DNA sequencing; missense variant; pathogenicity prediction

Address correspondence to: Meral Gunay-Aygun, M.D., NHGRI, NIH, 10 Center Dr. Bldg 10, Rm 10C103A, Bethesda, MD 20892, 301 594 4181, mgaygun@mail.nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Introduction

Autosomal recessive polycystic kidney disease (ARPKD), invariably associated with congenital hepatic fibrosis (CHF), is the most common childhood-onset ciliopathy, with an estimated frequency of 1 in 20,000 live births [1–5]. All typical ARPKD/CHF patients studied to date have been linked to chromosome 6p12, where *PKHD1*, the only gene mutated in ARPKD/CHF resides [6,7]. Clinically, ARPKD/CHF is characterized by non-obstructive dilatations of the renal collecting ducts resulting in progressive renal insufficiency and liver disease in the form of CHF and macroscopic biliary abnormalities [4,5]. Approximately half of ARPKD/CHF patients present in the perinatal period, with enlarged, echogenic kidneys and oligohydramnios, often leading to death secondary to pulmonary hypoplasia [1,2,8]. Most of the remaining patients present in childhood with kidney or liver related symptoms, and the minority of patients come to medical attention in adulthood with liver-related complications in association with mild kidney disease [9,10].

The diagnosis of ARPKD relies upon clinical findings, specifically radiographic abnormalities or biopsy evidence of typical renal or hepatic pathology [4,5]. Currently, DNA analysis of *PKHD1* is not part of routine clinical practice; it is used to confirm the diagnosis in difficult cases and for prenatal diagnosis [4,11]. This is in part due to the fact that *PKHD1* is a large and complicated gene. It spans approximately 470 kb of genomic DNA and consists of 86 exons variably assembled into a number of alternatively spliced transcripts ranging in size from 9 to 16 kb [6,12]. The mouse homologue of *PKHD1* also has a complex splicing patternsuggesting functional importance of the alternative spliced products. The longest open reading frame (ORF) of *PKHD1* is 12.2 kb in length and contains 67 exons that encode a 4074 amino acid protein called fibrocystin or polyductin (FCPD) [6,7]. FCPD is a novel receptor-like protein with a large extracellular, single transmembrane domain and a small intracytoplasmic domain. It contains multiple TIG/IPT domains (immunoglobin-like folds shared by plexins and transcription factors) and Parallel Beta-Helix 1 (PBH1) repeats. Some *PKHD1* transcripts that lack the transmembrane domain are predicted to be secreted if translated [6].

Since the identification of *PKHD1* in 2002 [6,7], several mutation detection studies have analyzed its longest ORF of *PKHD1*.[6–11,13–19] (Table 1). More than 300 pathogenic *PKHD1* variants dispersed throughout the gene are tabulated in a disease-specific DNA variation database (http://www.humgen.rwth-aachen.de/). Approximately 60% of the *PKHD1* pathogenic variants reported to date are truncating and 40% are missense mutations. A small number of relatively common mutations account for 10% – 20% of all *PKHD1* mutations [17]. The most common missense mutation in the *PKHD1* gene is c.107C>T (p.Thr36Met). This mutation is reported repeatedly in patient populations of various backgrounds and estimated to constitute 20% of all *PKHD1* mutations[17]. Other *PKHD1* mutations identified in more than one family include c.664A>G (p.Ile222Val), c.2414C>T (p.Pro805Leu), c.6992T>A (p.Ile2331Lys), c.8870T>C (p.Ile2957Thr), c.9530T>C (p.Ile3177Thr), c.10174C>T (p.Gln3392X), c.5895dupA (p.Leu1966fs), and c. 9689del A. (p.Asp3230fs), c.3761_3762del insG, (p.Ala1254fs); exact frequencies of these individual mutations are unknown. The remaining mutations are rare variants dispersed across the coding sequence of the gene. Approximately one third of *PKHD1* mutations are unique to a single family [20]. Some genotype-phenotype correlation exists; patients with 2 truncating mutations do not survive the neonatal complications. Survival beyond the newborn period requires the presence of at least one missense mutation [9]. The majority of the published cohorts are enriched with DNA samples from patients having the severe perinatal form of ARPKD (Table 1); most studies used a mutation screening method such as denaturing high-performance liquid chromatography (DHPLC) [6,7,9] or single-strand polymorphism analysis (SSCP) [14] (Table 1). Direct sequencing was performed in only one study [13].

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

In this study, we report direct sequencing results of the *PKHD1* gene on 78 patients from 68 families who fulfilled the clinical diagnostic criteria for ARPKD. Differing from previously published cohorts, our patient population was required to survive beyond 6 months of age, to travel to the NIH Clinical Center for evaluation and to have the diagnosis of ARPKD clinically confirmed. Here, we present our patients' novel and previously identified *PKHD1* variants, make comparisons with the published molecular and clinical data and discuss some of the challenges involved in interpreting the pathogenicity of missense variants in this large and complicated gene.

Methods

Clinical Assessments

The patients and their families were evaluated at the NIH Clinical Center under the intramural NIH protocol "Clinical Investigations into the Kidney and Liver Disease in Autosomal Recessive Polycystic Kidney Disease/Congenital Hepatic Fibrosis and other Ciliopathies" (www.clinicaltrials.gov, trial NCT00068224). Patients or their parents gave written, informed consent. Our cohort included 90 patients referred with a diagnosis of ARPKD. Evaluations at the NIH Clinical center included family history and physical examination by a pediatrician clinical geneticist (MGA) and comprehensive biochemical and imaging studies. Standard and high resolution ultrasonographic (HR-USG) studies were performed using 4 and 7 Mhz transducers (AVI Sequoia Inc, Mountain View, CA). Magnetic resonance imaging (MRI) and MR cholangiopancreatography (MRCP) were performed on 1.5 or 3 Tesla machines (Philips Medical Systems, NA, Bothell, Washington; General Electric Healthcare, Waukesha, WI, USA) without intravenous contrast media.

Seventy eight patients from 68 independent families, who fulfilled the established clinical diagnostic criteria [1,21] for ARPKD based upon their NIH evaluation, are included in this paper (Table 3). These clinical diagnostic criteria [1,21], included typical kidney and liver involvement on imaging and/or biopsy, absence of congenital malformations and autosomal recessive inheritance. Clinical features of the 12 patients who did not fulfill the clinical diagnostic criteria for ARPKD are listed in Table 2. Patients who were symptomatic at birth or up to day of life 30 were classified as perinatal presenters, and those who first became symptomatic after the first month of life were classified as nonperinatals. Patients diagnosed by prenatal USG were classified as nonperinatal if they remained asymptomatic during the first month of life. The families with multiple children with perinatal and non-perinatal presentations were classified in the perinatal group for mutation detection rate calculations. The parents who were available at the time of the NIH evaluation underwent screening abdominal ultrasound evaluations for renal and hepatic disease. In addition, parental blood samples were collected for DNA analysis for confirmation of segregation.

Molecular Studies

DNA was extracted from blood using Puregene kits (Germantown, MD). DNA sequencing in both sense and antisense directions was performed in our laboratory using a Beckman CEQ 8000 system and reagents (Beckman Coulter, Inc., Fullerton, CA) and by Agencourt BioScience (Beverly, MA) and ACGT, Inc. (Wheeling, IL). DNA sequencing was performed on all coding exons (2–67) of the longest ORF of *PKHD1* and their intronic boundaries which included on average 20–30 bp intronic sequence on both ends of the exons. These regions were amplified in 76 amplicons. Exons 32, 58, and 61 were sequenced in overlapping fragments due to their large size. PCR and DNA sequencing primers were initially taken from the existing literature [13]. Some primers were redesigned using the primer 3 program (http://frodo.wi.mit.edu/) and are available by request. Custom primer synthesis was carried out by Oligonet (Gaithersburg, MD). DNA alignment and sequence variant analysis were

NIH-PA Author Manuscript

carried out using Sequencher (GeneCodes, Ann Arbor, MI). Control DNA samples obtained from Coriell were screened for the identified novel missense mutations using the 5′ nuclease allelic discrimination (TaqMan) assay, as previously described [22] or by restriction fragment length polymorphism analysis. Reference sequences included genomic sequence from NC000006.10 and mRNA sequence from NM138694.

Pathogenicity Assessment

Several methods were used to evaluate the pathogenicity of the missense variants, although not all were used to assign the overall pathogenicity score. These included the following: 1. Consistency of segregation, checked by mutation analysis of the parents when available; 2. The general population frequencies of the novel missense variants, evaluated by analyzing 200 to 400 control chromosomes; 3. Missense variants, evaluated by 3 different web-based computational pathogenicity prediction tools, i.e., Align GVGD (http://agvgd.iarc.fr/agvgd_input.php), PolyPhen (http://coot.embl.de/PolyPhen/) and SNAP (http://cubic.bioc.columbia.edu/services/SNAP/); 4. Novel missense variants, evaluated by the splice variant interpretation software NetGene2 Server (http://www.cbs.dtu.dk/services/NetGene2/); 5. The *PKHD1*-specific mutation database (http://www.humgen.rwth-aachen.de/), the Human Genome Mutation Database (HGMD) (http://www.hgmd.cf.ac.uk/ac/index.php) and the previously published *PKHD1* mutation detection articles, reviewed for existing data about the variants.

Align GVGD is a Grantham matrix [23] based pathogenicity evaluation tool that uses multiple species' polypeptide alignments to determine the range of amino acid chemistries that can be tolerated at a specific amino acid position (Grantham variation, GV) and compares it to the magnitude of the difference between the wild type and the identified amino acid change (Grantham deviation, GD). Align GVGD classifies variants into 7 groups (class 0, class 15, class 25, class 35, class 45, class 55 and class 65) ranging from least likely (class 0) to most likely (class 65) to interfere with the function of the protein. For Align GVGD analysis, we used two different multispecies alignments. For Align GVGD prediction 1, we aligned *Homo sapiens* fibrocystin with its homologues in *Pan troglodytes (chimp), Mus musculus (mouse), Canis lupis familiaris (dog),* and *Gallus gallus (chicken)*. For Align GVGD prediction 2, we aligned only human and mouse. These alignments were constructed using HomoloGene (http://www.ncbi.nlm.nih.gov/homologene).

The PolyPhen (Polymorphism Phenotyping) computational pathogenicity prediction tool combines several types of analysis including experimentally-determined structure (if available), analytically determined structure (based on local amino acid sequence), and multiple sequence alignments [24–27]. PolyPhen analysis classifies missense variants into 3 groups as benign, possibly pathogenic and probably pathogenic.

SNAP (Screening for Non-Acceptable Polymorphisms) is another computational missense variant evaluation tool that uses several sources of information, including alignment to related protein motifs, secondary structure predictions and solvent accessibility calculations based on predicted structure to determine whether a polymorphism is likely to be either neutral or non-neutral [28,29]. One advantage of SNAP is that it provides an estimate for the accuracy of the prediction.

Combining the above methods, we assigned an "overall pathogenicity score" of (1) to pathogenic mutations, which we defined as those due to protein truncating mutations caused by either nonsense variants or out-of-frame in/dels. For missense sequence variants, we used the following criteria: "Probably pathogenic" (2), not identified in at least 200 control chromosomes, Align GVGD prediction 1 (based on 5-species multialignment) equal to or higher than class 35 in combination with a non-benign PolyPhen or a non-neutral SNAP

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

prediction; “Possibly pathogenic” (3), not identified in at least 200 control chromosomes, Grantham prediction 1 lower than class 35, Grantham prediction 2 equal to or higher than class 25 in combination with non-benign PolyPhen or a non-neutral SNAP predictions; “Probably benign” (4), missense variants that did not meet the above criteria and variants previously reported as polymorphisms. These same criteria were applied to missense variants that were previously reported as pathogenic only once. The missense variants repeatedly reported as pathogenic were assigned an overall pathogenicity estimate of “probably pathogenic” (2), independent of the Align GVGD, PolyPhen and SNAP predictions.

Results

Upon evaluation of the 90 probable ARPKD patients at the NIH Clinical Center, the clinical diagnosis of ARPKD was confirmed in 78 patients from 68 independent families (Table 3). One family (#43) contributed an aunt and niece pair, 1 family (#10) had 4 affected siblings and 6 families contributed 2 affected siblings each. Table 3 lists the ethnic background, sex and age at diagnosis, and individuals with CHF-predominant disease, as well as age at onset of symptoms; 33 of 68 families (49 %) were classified as perinatal and 35 (51%) as nonperinatal. Six of the 68 families (9%) were classified as perinatally-fatal since they contained at least 1 child who did not survive the perinatal complications. Thirteen individuals from 12 families (18%) were classified as CHF-predominant based on severe CHF-related manifestations in association with mild kidney disease.

The DNA of the 78 clinically confirmed ARPKD patients was sequenced for the 67 coding exons (exons 2–67) of the longest ORF of *PKHD1*. There were 137 independent patient alleles from 68 families because of the presence of 3 independent alleles in the aunt-niece family.

In total, 77 *PKHD1* sequence variants were identified; 41 (53%) were novel and 36 were previously described (Tables 3 and 4). Nineteen of the 77 variants were truncating mutations caused by either nonsense alterations or frameshifting small in/dels. Nine of the 19 truncating mutations were novel; 10 were previously described. Of the 55 missense variants, 31 were novel and 24 were previously described. The remaining 3 variants were a novel in-frame deletion of one amino acid and 2 previously described canonical splice site mutations one of which (c.8642+1G>A) was previously reported in the same patient [30].

We combined multiple approaches to assess the pathogenicity of the 55 missense variants (Table 4). These approaches included analysis of 200 to 400 control chromosomes to determine the population frequencies, use of 3 different computational pathogenicity prediction tools (Align GVGD, PolyPhen and SNAP), and evaluation of the potential splicing effects of the novel missense variants by using splice variant identification software NetGene2. Based on the results of these evaluations, and the criteria described in the Methods section, the 55 missense variants were classified into four pathogenicity estimate groups as follows: 1. Pathogenic; 2. Probably pathogenic; 3. Possibly pathogenic; and 4. Probably benign (Table 4, “overall pathogenicity score” column). Of the 31 novel missense variants, 19 were estimated to be probably pathogenic, 8 possibly pathogenic and 4 probably benign. Of the 24 previously reported missense variants, 17 were classified as probably pathogenic, 2 possibly pathogenic and 5 probably benign.

Considering the sequence variants with pathogenicity scores 1, 2 or 3, the overall mutation detection rate in the present study was 79 % (108/137). At least 2 variants with pathogenicity scores 1, 2 or 3 were detected in 44 families, one was identified in 20 families and no pathogenic variants were found in 5 patients (Table 3).

The distribution of the *PKHD1* sequence variants among patients is listed in Table 3. No families having 2 truncating mutations (either frameshift or nonsense) were identified in the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

present cohort. Truncating mutations were identified in 23 families, with 18 of these being in combination with a missense variant. The remaining 40 families had the following combination of variants: 23 missense variants on both alleles, 14 with only one missense variant, 2 with one missense variant and one splice variant and 1 with a missense variant in combination with a single amino acid deletion. In 5 families, no sequence variants were identified. When we compared the perinatal and nonperinatal ARPKD patient groups, the frequency of missense variants that result in a change in the chemical class of the amino acid was comparable. Among perinatal onset ARPKD patients, 76% (35 of 46) of missense variants resulted in a change of the chemical class of the amino acid; and among the nonperinatal ARPKD group, this figure was 79% (45 of 57).

The disease manifested a perinatal onset in 33 families, 6 of whom experienced perinatally fatal ARPKD in another sibling. The mutation detection rate for the 33 perinatal families was 80% (53/66); 10 truncating mutations were identified in this group. The mutation detection rate for the 35 nonperinatal families was 77% (54/70) with 13 truncating mutations. When families were divided into “CHF-predominant” and others, mutation detection rate for the CHF-predominant group was 63% (15/24), while the detection rate for the remainder of the group was 82% (92/112).

In 8 of the 68 families, more than 2 sequence variants were identified; in 5 of these families more than 2 variants were classified as pathogenicity score 1–3 (Table 3). In 3 of these 5 families we were able to determine the parental inheritance phase of these variants. In family #1, p.Thr36Met was inherited from the mother and p.Asp3230fs was on the paternal chromosome; p.Tyr2661His was not found in either parent and therefore was thought to be a *de novo* change in this patient. In family #20, p.Arg781X and p.Arg3957Cys mutations were inherited on the same maternal chromosome, while p.His3049Arg was inherited paternally. In family #30, p.Ile307Thr was maternal in origin. Therefore, it was inferred from this result that p.Gly2705fs and p.Ser2861Gly mutations were on the same allele but paternal DNA was not available. In family number #33, p.Leu2106Arg and p.Val3219Ala were inherited on the paternal chromosome and p.Ile292Val was inherited from the mother. In family #34, p.Tyr486X and p.Tyr1136Cys were on the same paternal chromosome and p.Ile246Thr was inherited from the mother.

Discussion

PKHD1 is one of the largest and most complicated genes in the human genome. The disease spectrum caused by mutations in *PKHD1* is similarly complex, ranging from perinatally-fatal PKD to CHF-predominant presentations in adulthood with mild or no apparent kidney disease. Despite these challenges, several large and informative *PKHD1* mutation detection studies have been published [6,7,9,10,[13],14,16,19,31]. Major characteristics of these and the present study are summarized in Table 1. Due to the large size of the gene, DHPLC or SSCP screening techniques have been used in all but one of the published studies; variants detected by screening were further characterized by targeted direct sequencing. Given the wide range of the ARPKD/CHF phenotypes, these cohorts naturally contained different proportions of samples from patients with severe (perinatally-fatal) and relatively milder forms of ARPKD/CHF. The percentage of the perinatally-fatal patients in the published cohorts ranged from 10% to 100 % (Table 1). Methods used to evaluate the pathogenicity of missense mutations also varied. While more traditional strategies (confirmation of segregation, detection of the variant’s frequency in the general population, evaluation of conservation in the mouse homolog and the magnitude of the chemical change caused by the new amino acid) were used in most studies, other methods (web-based computational pathogenicity prediction tools) were used in some. The number of control chromosomes analyzed varied between 100 and 400.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

The overall mutation detection rate in the published cohorts ranged from 42% to 87% (Table 1). The mutation detection rate of the present study is 79%. Several characteristics of the present investigation are different from those of previously published studies. Our cohort was relatively enriched in later-onset ARPKD/CHF; the proportion of perinatally-fatal ARPKD was 9%, the lowest among the series reported to date (Table 1). This is due to the fact that enrollment in our cohort required survival of at least one affected family member (the patient examined at NIH) beyond 6 months of age and the ability to travel to the NIH. In addition, our cohort encompassed a wide age range (1 to 56 years) of patients including older children and adults with a CHF-predominant phenotype. The data show higher mutation detection rates in the perinatally symptomatic ARPKD cohorts because such patients are more likely to have protein truncating mutations that are relatively easy to detect. This is exemplified in the relatively lower mutation detection rates reported by Bergmann et al. (2003) in non-perinatally fatal patients (40%) in comparison to that in the perinatally-fatal group (77%) (Table 1). Similarly, Furu et al. (2003) reported mutation detection rates of 32%, 42% and 85% in CHF, non-perinatally fatal ARPKD and perinatally-fatal ARPKD, respectively (Table 1). Consistently, the relative percentages of truncating and missense mutations in the reported series varied between 18 and 65%, largely reflecting the percentages of the perinatally-fatal patients in each cohort. In our cohort, 28% of the potentially pathogenic variants were truncating. Similar to Bergmann et al.'s (2003) and Furu et al.'s (2003) results, our mutation detection rate among CHF-predominant patients (63%) was lower than that for the remainder of the cohort (82%). Given the relatively low representation of the severe perinatally-fatal form of ARPKD patients in the present cohort, our overall mutation detection rate of 79% is at the higher end of the expected range.

No mutations were identified in 5 (7%) of our 68 independent families. Both mutations of these patients may be difficult to identify, perhaps because they reside in parts of the gene we have not sequenced (deep intronic, 3′ and 5′ UTRs, non coding exons, promoter and other regulatory regions) or because they could not be detected by direct sequencing (large deletions/ rearrangements). Alternatively, these patients might represent phenocopies of ARPKD. However, inclusion in the present cohort required patients to undergo extensive evaluations at the NIH for confirmation of the clinical diagnosis of ARPKD, including a detailed family history and physical examination, USG and MRI imaging, and biochemical testing. In fact, 12 of the 90 patients who carried a diagnosis of ARPKD upon admission to the NIH Clinical Center were not included in the presented cohort. This decreases the likelihood of existence of easily recognizable phenocopies of ARPKD (such as Bardet-Biedl, Oral-facial-digital or Joubert syndromes and related ciliopathies) in our cohort. However, the presence of closer phenocopies, perhaps undistinguisable by imaging and potentially even by histopathology, remains possible.

Our data support previously published genotype-phenotype correlation findings. Consistent with the previous observation that survival beyond the newborn period requires the presence of at least 1 missense mutation, [15] no patients with 2 truncating mutations were identified in our cohort. The missense mutations p.Tyr486His, p.Pro805Leu, p.Ile3177Thr, p.Cys1472Tyr, p.Ile2303Phe, p.His3124Tyr, Leu2134Pro and Asp2761Tyr and p.Arg3482Cys were previously reported to be associated with a severe perinatally-fatal phenotype [8,14]. Consistent with this observation, none of these mutations was identified in our cohort. Fibrocystin has a very large extracellular domain (amino acids 1 – 3858), one transmembrane domain (amino acids 3859 – 3879) and a small intracellular domain (amino acids 3880 – 4074). Consistent with prior reports, the 41 novel *PKHD1* variants we identified in this study were dispersed throughout the fibrocystin protein without clustering at specific domains. All previously reported missense variants, and all but 1 of the 31 novel missense variants identified in this study reside in the extracellular domain of fibrocystin between amino acids 36 and 3219; 1 novel missense variant (p.Arg3957Cys) lies on the intracellular domain. No mutation was

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

found in the transmembrane domain. All truncating mutations identified in this study resulted in premature stop codons upstream of the transmembrane region. Within the extracellular domain, mutations were distributed randomly, without any concentration in the known domains of the protein, including the multiple immunoglobulin like plexin-transcription factor domains and the parallel β-helix 1 repeats.

Deciding whether a novel missense variant is disease-causing or harmless is a challenge, especially in the absence of a reliable functional assay or the crystallized structure of the protein (FCPD) in question. In the case of pathogenic missense variants that are not extremely rare, recurrent detection of the same variant in patient populations, disproportionate to its frequency in the general population, supports pathogenicity. However, this “test of time” might not always be helpful, especially for very large genes with many rare mutations dispersed throughout all coding exons, as for *PKHD1,* because a truly pathogenic mutation might be identified only once even when large patient cohorts are combined. In an effort to maximize the accuracy of our pathogenicity estimates for the missense variants, we combined various methods. In addition to determining the frequency of the variant in the general population, we used 3 web-based computational missense variant pathogenicity prediction tools and a splice variant prediction tool. In the absence of crystallized protein structure, the predictions made by computational tools such as Align GVGD, PolyPhen, and SNAP depend largely upon multiple species homologue sequence alignment for evaluation of evolutionary tolerance to variance in a given amino acid. The larger the evolutionary distance between the aligned species, the lower the risk of overpredicting pathogenicity. Align GVGD determines the range of tolerance to variation for the amino acid position in question by aligning the homologues of the protein from various species, and compares this tolerance to the magnitude of the chemical difference caused by the detected variant. Align GVGD prediction 1, based on an alignment comparing 5 species, is more stringent than Align GVGD prediction 2, which compares only human and mouse. Some tools such as PolyPhen and SNAP also integrate input from the neighboring sequences (whether an important protein domain or not) and from the predicted structure of the protein. The predictions made by these computational tools for our cohort’s missense variants are listed in Table 4.

Many of the methods used for deciding about the pathogenicity of missense variants have inherent limitations. Segregation analysis suggests that a given missense variant is more likely to be a polymorphism if the variant in question is on the same chromosome with a truncating variant. However, consistent segregation of two variants in the family (one from each parent) does not always mean pathogenicity because harmless variants might also be inherited one from each parent. The population frequency of a given variant of unknown significance is helpful only when its frequency is inordinately high in comparison to the expected frequency of a given mutation in that specific gene, based on the observed frequency of the disease. When a variant is very rare and not identified in 200 chromosomes, this favors pathogenicity but it can still be a very rare harmless variant. The false negative and false positive prediction rates of the computational prediction tools are 10%–20% [32]. Estimation of error-rates may be complicated by the presence of misclassified variants in the datasets used to design and test these softwares. Several examples illustrating the limitations of the pathogenicity prediction tools are listed in the “previously reported missense variants” section of Table 4.

We identified 5 families with more than 2 variants assigned to be potentially pathogenic (scores 1–3). In 3 of these 5 families we were able to determine the “cis” or “trans” status of these variants by determining the phase in parental chromosomes. In Family #20, p.Arg3957Cys was on the same chromosome with truncating mutation p.Arg781X, making it unlikely that it contributed to the clinical phenotype, at least in this family. Similarly, missense mutations p.Ser2861Gly and p.Tyr1136Cys in families #30 and #34, respectively, were on the same chromosome with a truncating mutation, making it unlikely that these missense variants

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

contributed to the phenotype, at least in these families. Since it is theoretically possible that these 3 missense variants can still be pathogenic when present by themselves, we did not change their overall pathogenicity predictions.

Given the limitations of the variable pathogenicity assessments methods, it is conceivable that some of the missense variants are misclassified in the present study or in previously published reports. It remains to be determined whether some “hypomorphic” missense variants might contribute to the phenotype when in combination with certain severe mutations. Some of these questions may be answered when the structural and functional characteristics of fibrocystin are better defined. *PKHD1* sequencing, preferably including promoter and other regulatory regions, in more ARPKD patients might allow better classification of missense variants and increase the overall mutation detection rate. In the meantime, use of *PKHD1* sequencing data for important clinical decisions such as prenatal diagnosis will require caution, especially when novel or rarely reported missense variants are the only mutations identified in a given family.

Acknowledgments

We thank the ARPKD/CHF Alliance for their extensive support of this protocol and the patients and their families who generously participated in this investigation. Supported by the Intramural Research Programs of the National Human Genome Research Institute, National Cancer Institute, National Institute of Diabetes and Digestive and Kidney Diseases and the NIH Clinical Center.

References

1. Guay-Woodford LM, Desmond RA. Autosomal recessive polycystic kidney disease: the clinical experience in North America. Pediatrics 2003;111:1072–1080. [PubMed: 12728091]
2. Zerres K, Rudnik-Schoneborn S, Deget F, Holtkamp U, Brodehl J, Geisert J, Scharer K. Autosomal recessive polycystic kidney disease in 115 children: clinical presentation, course and influence of gender. Arbeitsgemeinschaft fur Padiatrische, Nephrologie Acta Paediatr 1996;85:437–445.
3. Gunay-Aygun M, Avner ED, Bacallao RL, Choyke PL, Flynn JT, Germino GG, Guay-Woodford L, Harris P, Heller T, Ingelfinger J, Kaskel F, Kleta R, LaRusso NF, Mohan P, Pazour GJ, Shneider BL, Torres VE, Wilson P, Zak C, Zhou J, Gahl WA. Autosomal recessive polycystic kidney disease and congenital hepatic fibrosis: summary statement of a first National Institutes of Health/Office of Rare Diseases conference. The Journal of pediatrics 2006;149:159–164. [PubMed: 16887426]
4. Dell, KM.; Avner, ED. GeneReviews at GeneTests: Medical Genetics Information Resource (database online). Copyright, University of Washington; Seattle: 1997–2008. Autosomal recessive polycystic kidney disease. Available at http://www.genetests.org., 2008
5. Gunay-Aygun, MG.; Heller, TWA. GeneReviews at GeneTests: Medical Genetics Information Resource (database online). Copyright, University of Washington; Seattle: 1997–2008. Congenital hepatic fibrosis overview. Available at http://www.genetests.org. (2008)
6. Onuchic LF, Furu L, Nagasawa Y, Hou X, Eggermann T, Ren Z, Bergmann C, Senderek J, Esquivel E, Zeltner R, Rudnik-Schoneborn S, Mrug M, Sweeney W, Avner ED, Zerres K, Guay-Woodford LM, Somlo S, Germino GG. PKHD1, the polycystic kidney and hepatic disease 1 gene, encodes a novel large protein containing multiple immunoglobulin-like plexin-transcription-factor domains and parallel beta-helix 1 repeats. Am J Hum Genet 2002;70:1305–1317. [PubMed: 11898128]
7. Ward CJ, Hogan MC, Rossetti S, Walker D, Sneddon T, Wang X, Kubly V, Cunningham JM, Bacallao R, Ishibashi M, Milliner DS, Torres VE, Harris PC. The gene mutated in autosomal recessive polycystic kidney disease encodes a large, receptor-like protein. Nature genetics 2002;30:259–269. [PubMed: 11919560]
8. Bergmann C, Senderek J, Windelen E, Kupper F, Middeldorf I, Schneider F, Dornia C, Rudnik-Schoneborn S, Konrad M, Schmitt CP, Seeman T, Neuhaus TJ, Vester U, Kirfel J, Buttner R, Zerres K. Clinical consequences of PKHD1 mutations in 164 patients with autosomal-recessive polycystic kidney disease (ARPKD). Kidney international 2005;67:829–848. [PubMed: 15698423]
9. Furu L, Onuchic LF, Gharavi A, Hou X, Esquivel EL, Nagasawa Y, Bergmann C, Senderek J, Avner E, Zerres K, Germino GG, Guay-Woodford LM, Somlo S. Milder presentation of recessive polycystic

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

kidney disease requires presence of amino acid substitution mutations. J Am Soc Nephrol 2003;14:2004–2014. [PubMed: 12874454]

10. Adeva M, El-Youssef M, Rossetti S, Kamath PS, Kubly V, Consugar MB, Milliner DM, King BF, Torres VE, Harris PC. Clinical and molecular characterization defines a broadened spectrum of autosomal recessive polycystic kidney disease (ARPKD). Medicine 2006;85:1–21. [PubMed: 16523049]
11. Zerres K, Senderek J, Rudnik-Schoneborn S, Eggermann T, Kunze J, Mononen T, Kaariainen H, Kirfel J, Moser M, Buettner R, Bergmann C. New options for prenatal diagnosis in autosomal recessive polycystic kidney disease by mutation analysis of the PKHD1 gene. Clinical genetics 2004;66:53–57. [PubMed: 15200508]
12. Harris P. Molecular basis of polycystic kidney disease: PKD1, PKD2 and PKHD1. Current opinion in nephrology and hypertension 2002;11:309–314. [PubMed: 11981261]
13. Losekoot M, Haarloo C, Ruivenkamp C, White SJ, Breuning MH, Peters DJ. Analysis of missense variants in the PKHD1-gene in patients with autosomal recessive polycystic kidney disease (ARPKD). Human genetics 2005;118:185–206. [PubMed: 16133180]
14. Bergmann C, Senderek J, Sedlacek B, Pegiazoglou I, Puglia P, Eggermann T, Rudnik-Schoneborn S, Furu L, Onuchic LF, De Baca M, Germino GG, Guay-Woodford L, Somlo S, Moser M, Buttner R, Zerres K. Spectrum of mutations in the gene for autosomal recessive polycystic kidney disease (ARPKD/PKHD1). J Am Soc Nephrol 2003;14:76–89. [PubMed: 12506140]
15. Bergmann C, Senderek J, Schneider F, Dornia C, Kupper F, Eggermann T, Rudnik-Schoneborn S, Kirfel J, Moser M, Buttner R, Zerres K. PKHD1 mutations in families requesting prenatal diagnosis for autosomal recessive polycystic kidney disease (ARPKD). Hum Mutat 2004;23:487–495. [PubMed: 15108281]
16. Rossetti S, Torra R, Coto E, Consugar M, Kubly V, Malaga S, Navarro M, El-Youssef M, Torres VE, Harris PC. A complete mutation screen of PKHD1 in autosomal-recessive polycystic kidney disease (ARPKD) pedigrees. Kidney international 2003;64:391–403. [PubMed: 12846734]
17. Bergmann C, Kupper F, Dornia C, Schneider F, Senderek J, Zerres K. Algorithm for efficient PKHD1 mutation screening in autosomal recessive polycystic kidney disease (ARPKD). Human mutation 2005;25:225–231. [PubMed: 15706593]
18. Bergmann C, Kupper F, Schmitt CP, Vester U, Neuhaus TJ, Senderek J, Zerres K. Multi-exon deletions of the PKHD1 gene cause autosomal recessive polycystic kidney disease (ARPKD). Journal of medical genetics 2005;42:e63. [PubMed: 16199545]
19. Sharp AM, Messiaen LM, Page G, Antignac C, Gubler MC, Onuchic LF, Somlo S, Germino GG, Guay-Woodford LM. Comprehensive genomic analysis of PKHD1 mutations in ARPKD cohorts. Journal of medical genetics 2005;42:336–349. [PubMed: 15805161]
20. Rossetti S, Harris PC. Genotype-phenotype correlations in autosomal dominant and autosomal recessive polycystic kidney disease. J Am Soc Nephrol 2007;18:1374–1380. [PubMed: 17429049]
21. Zerres K, Volpel MC, Weiss H. Cystic kidneys. Genetics, pathologic anatomy, clinical picture, and prenatal diagnosis. Hum Genet 1984;68:104–135. [PubMed: 6500563]
22. Livak KJ. Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. Genet Anal 1999;14:143–149. [PubMed: 10084106]
23. Miller MP, Kumar S. Understanding human disease mutations through the use of interspecific genetic variation. Hum Mol Genet 2001;10:2319–2328. [PubMed: 11689479]
24. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 2002;30:3894–3900. [PubMed: 12202775]
25. Sunyaev S, Ramensky V, Bork P. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. Trends Genet 2000;16:198–200. [PubMed: 10782110]
26. Sunyaev S, Ramensky V, Koch I, Lathe W 3rd, Kondrashov AS, Bork P. Prediction of deleterious human alleles. Hum Mol Genet 2001;10:591–597. [PubMed: 11230178]
27. Sunyaev SR, Lathe WC 3rd, Ramensky VE, Bork P. SNP frequencies in human genes an excess of rare alleles and differing modes of selection. Trends Genet 2000;16:335–337. [PubMed: 10904261]
28. Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res 2007;35:3823–3835. [PubMed: 17526529]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

29. Bromberg Y, Yachdav G, Rost B. SNAP predicts effect of mutations on protein function. Bioinformatics 2008;24:2397–2398. [PubMed: 18757876]
30. Sgro M, Rossetti S, Barozzino T, Toi A, Langer J, Harris PC, Harvey E, Chitayat D. Caroli's disease: prenatal diagnosis, postnatal outcome and genetic analysis. Ultrasound Obstet Gynecol 2004;23:73–76. [PubMed: 14971004]
31. Bergmann C, Senderek J, Kupper F, Schneider F, Dornia C, Windelen E, Eggermann T, Rudnik-Schoneborn S, Kirfel J, Furu L, Onuchic LF, Rossetti S, Harris PC, Somlo S, Guay-Woodford L, Germino GG, Moser M, Buttner R, Zerres K. PKHD1 mutations in autosomal recessive polycystic kidney disease (ARPKD). Human mutation 2004;23:453–463. [PubMed: 15108277]
32. Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. Annu Rev Genomics Hum Genet 2006;7:61–80. [PubMed: 16824020]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

Summary of *PKHD1* sequence analysis studies.

PKHD1 DNA analysis studies	Method	Species used for conservation analysis	Computational pathogenicity prediction tools	Number of control chromosomes analyzed	Number of independent ARPKD/ CHF chromosomes sequenced	Percentage of perinatally-fatal ARPKD	Mutation detection rate	Truncating (%)	Missense (%)	Splice (%)	Large *PKHD1* deletions	Alternative *PKHD1* exons
Present Study	Direct sequencing	Align GVGD (Human, chimp, mouse, dog, chicken) PolyPhen (human, mouse, rat, dog)	Align GVGD, PolyPhen, SNAP	200 – 400	137*	9%	79%	28%	68%	3%	Not evaluated	Not evaluated
Adeva et al., 2006	DHPLC	Human, mouse, rat, chicken	Not reported	200	62	10%	76%	21%	77%	2%	Not evaluated	Not evaluated
Sharp et al., 2005	DHPLC	Human, mouse, rat, +/− chicken	Grantham matrix based evaluation using criteria by Miller and Kumar 2001	200	150**	65%	83%	18%	69%	13%	Not evaluated	19 alternative exons sequenced (data not shown)
Bergmann et al., 2005b	DHPLC; RT-PCR	Not applicable	Not applicable	Not applicable	116***	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	3 partial gene deletions	19 alternative exons sequenced. Several missense variants of unknown significance identified
Losekoot et al., 2005	Direct sequencing; MLPA	Human, chimp, dog, mouse, frog	Grantham matrix based evaluation using criteria by Abkevich et al. 2004	Not evaluated	78	65%	87%	38%	55%	7%	None detected	Alternative exons 38, 62, 63 and 64 sequenced in families with 1 mutation only; no mutations identified
Bergmann et al., 2004b	DHPLC	Human and mouse	Not reported	400	80	100%	85%	65%	33%	2%	Not evaluated	Not evaluated
Bergmann et al., 2003	SSCP	Human and mouse	Not reported	300	180	49%	Overall: 61% Perinatally-fatal: 77%, Perinatally non-fatal: 40%	43%	55%	2%	Not evaluated	Not evaluated
Furu et al., 2003	DHPLC	Not reported	Not reported	320	120	NA	Overall 52% Perinatally-fatal 85%, Moderate ARPKD 42% CHF 32%	39%	61%	0%	Not evaluated	Not evaluated
Rosetti et al., 2003	DHPLC	Human and mouse	Not reported	100	122	17%	47%	24%	64%	12%	Not evaluated	Not evaluated
Onuchic et al., 2002	DHPLC	Human and mouse	Not reported	120	50	NA	42%	53%	47%	0%	Not evaluated	Not evaluated

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

PKHD1 DNA analysis studies	Method	Species used for conservation analysis	Computational pathogenicity prediction tools	Number of control chromosomes analyzed	Number of independent ARPKD/ CHF chromosomes sequenced	Percentage of perinatally-fatal ARPKD	Mutation detection rate	Truncating (%)	Missense (%)	Splice (%)	Large *PKHD1* deletions	Alternative *PKHD1* exons
Ward et al., 2002	DHPLC; Southern blot	Human and mouse	Not reported	200	28	NA	68%	32%	67%	0%	None detected	Not evaluated

DHPLC: Denaturing high-performance liquid chromatography; RT-PCR: Real time-polymerase chain reaction; MLPA: Multiplex ligation-dependent probe amplification. SSCP: single-strand polymorphism analysis.

*This is an odd number because of a niece and aunt in the same family sharing 1 chromosome.

§ Included 12 alleles previously reported by Furu et al.

** 55 of these alleles had a known point mutation in one of the 66 ORF exons.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Features of the 12 patients who did not fulfill the clinical diagnostic criteria for ARPKD

Age (y)	Sex	Features not consistent with ARPKD	Diagnosis
1.8	M	No convincing imaging or laboratory evidence for congenital hepatic fibrosis. Renal ultrasound not consistent with ARPKD; multiple small (<1 cm) round macrocysts confined to the cortex	Probable glomerulocystic kidney disease
4.5	M	Facial dysmorphism, otherwise kidney and liver findings typical for ARPKD	Unknown
6	F	Renal ultrasound not consistent with ARPKD; multiple angiomyolipoma-like solid masses in addition to cysts (no other features of tuberosclerosis)	Unknown
6	F	Renal ultrasound not consistent with ARPKD; multiple angiomyolipoma-like solid masses in addition to cysts (no other features of tuberosclerosis)	Unknown
6	M	Dandy-Walker malformation, otherwise kidney and liver findings typical for ARPKD	Unknown
7	M	Renal ultrasound not consistent with ARPKD; multiple small (<1 cm) round macrocysts confined to the cortex	Probable glomerulocystic kidney disease
8	M	Developmental delay, otherwise kidney and liver findings typical for ARPKD	MKS3*-related ciliopathy (molecularly confirmed)
13	F	Renal ultrasound not consistent with ARPKD; macrocysts lining corticomedullary junction	Possible nephronophthisis
16	M	No convincing imaging or laboratory evidence for kidney involvement	Unknown
16	M	No convincing imaging or laboratory evidence for congenital hepatic fibrosis	Possible ADPKD
28	F	No convincing imaging or laboratory evidence for congenital hepatic fibrosis	Unknown
29	F	Facial dysmorphism, otherwise kidney and liver findings typical for ARPKD	Unknown

ADPKD: Autosomal dominant polycystic kidney disease;

*_MKS3_ gene was originally identified as one of the genes that cause Meckel syndrome.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 3

jor characteristics of the ARPKD patients enrolled in the present study and their *PKHD1* sequence variants.

mily No	Patient No	Sex	Ethnic Origin	Age at diagnosis (y)	Presentation	CHF predominant	Exon	Genomic DNA	*PKHD1* variants Coding DNA	Protein	Pathogenicity score
1	**1**	M	Caucasian	22 w	Perinatal Perinatal, sibling death	−	58	g.337007delA	c.9689delA	p.Asp3230fs	1
							3	g.1733C>T	c.107C>T	p.Thr36Met	2
							50	g.237033T>C	c.7981T>C	p.Tyr2661His	4
2	**2**	M	Caucasian	22 w	Perinatal Perinatal,sibling death	−	45	g.198973delT	c.7120delT	p.Phe2374fs	1
							16	g.26508G>T	c.1409G>T	p.Gly470Val	2
3	**3**	F	Caucasian	23 w	Perinatal	−	61	g.425261dupT	c.10452dupT	p.Phe3485fs	1
4	**4**	M	Caucasian	23 w	Perinatal	−	50	g.237120T>C	c.8068T>C	p.Trp2690Arg	2
							IVS55	-	c.8642+1G>A	-	2
5	**5**	F	Caucasian	28 w	Perinatal Perinatal,sibling death	−	48	g.217055C>T	c.7717C>T	p.Arg2573Cys	2
6	**6**	F	Hispanic	28 w	Perinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							46	g.201755T>C	c.7264T>C	p.Cys2422Arg	2
7	**7.1**	M	Caucasian	29 w	Perinatal	−	9	g.13925A>G	c.664A>G	p.Ile222Val	2
							58	g.336369_336371delATC	c.9048_9050delATC	p.Ser3017del	2
7	**7.2**	M	Caucasian	29w	Perinatal	−	9	g.13925A>G	c.664A>G	p.Ile222Val	2
							58	g.336369_336371delATC	c.9048_9050delATC	p.Ser3017del	2
8	**8**	F	Caucasian	29 w	Perinatal	−	22	g.34669C>G	c.2171C>G	p.Pro724Arg	3
							58	g.336464A>G	c.9146A>G	p.His3049Arg	3
9	**9**	M	Hispanic	29 w	Perinatal	−	53	g.293665T>C	c.8407T>C	p.Cys2803Arg	2
							53	g.293651T>G	c.8393T>G	p.Val2798Gly	2
10	**10.1**	F	Caucasian	29 w	Perinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							9	g.13925A>G	c.664A>G	p.Ile222Val	2
10	**10.2**	F	Caucasian	2	Nonperinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							9	g.13925A>G	c.664A>G	p.Ile222Val	2
10	**10.3**	F	Caucasian	5	Nonperinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							9	g.13925A>G	c.664A>G	p.Ile222Val	2
10	**10.4**	M	Caucasian	9	Nonperinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

mily No	Patient No	Sex	Ethnic Origin	Age at diagnosis (y)	Presentation	CHF predominant	Exon	Genomic DNA	*PKHD1* variants Coding DNA	Protein	Pathogenicity score
							9	g.13925A>G	c.664A>G	p.Ile222Val	2
11	**11**	F	Caucasian	30 w	Perinatal Perinatal,sibling death	−	52	g.254017_254018delGGinsCC	c.8246_8247delGGinsCC	p.Trp2749Ser	2
							IVS39	-	c.6490+2T>C	-	2
12	**12**	F	Caucasian	30 w	Perinatal Perinatal,sibling death	−	61	g.425445delT	c.10637delT	p.Val3546fs	1
							57	g.331653T>C	c.8870T>C	p.Ile2957Thr	2
13	**13**	M	Caucasian	30 w	Perinatal	−	51	g.248470delG	c.8114delG	p.Gly2705fs	1
							3	g.1733C>T	c.107C>T	p.Thr36Met	2
							55	g.312171A>G	c.8581A>G	p.Ser2861Gly	2
14	**14**	M	Caucasian	31 w	Perinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							3	g.1733C>T	c.107C>T	p.Thr36Met	2
15	**15**	F	Caucasian	38 w	Perinatal	−	61	g.425436_425444del8	c.10628_10635del8	p.Leu3543fs	1
							9	g.13925A>G	c.664A>G	p.Ile222Val	2
16	**16**	F	Caucasian	38 w	Perinatal Perinatal,sibling death	−	30	g.56685C>T	c.3467C>T	p.Ser1156Leu	2
							53	g.293669T>A	c.8410T>A	p.Met2804Lys	2
17	**17**	F	Caucasian	0	Perinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							43	g.181333T>A	c.6992T>A	p.Ile2331Lys	2
18	**18**	M	Caucasian	0	Perinatal	−	32	g.58890delC	c.3766delC	p.Gln1256fs	1
19	**19**	M	Caucasian	0	Perinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							36	g.124939G>T	c.5783G>T	p.Trp1928Leu	2
20	**20**	M	Caucasian	0	Perinatal	−	23	g.36376C>T	c.2341C>T	p.Arg781X	1
							67	g.465497C>T	c.11869C>T	p.Arg3957Cys	2
							58	g.336464A>G	c.9146A>G	p.His3049Arg	2
21	**21**	F	Caucasian	0	Perinatal	−	-	-	-	-	-
22	**22**	F	Caucasian	0	Perinatal	−	-	-	-	-	-
23	**23**	M	Caucasian	0	Perinatal	−	16	g.26496G>A	c.1397G>A	p.Gly466Glu	2
24	**24**	F	Caucasian	0	Perinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							3	g.1733C>T	c.107C>T	p.Thr36Met	2
25	**25**	F	Caucasian	0	Perinatal	−	16	g.26496G>A	c.1397G>A	p.Gly466Glu	2
							34	g.67374T>G	c.5450T>G	p.Val1817Gly	2

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

mily No	Patient No	Sex	Ethnic Origin	Age at diagnosis (y)	Presentation	CHF predominant	Exon	Genomic DNA	*PKHD1* variants Coding DNA	Protein	Pathogenicity score
							61	g.425734G>A	c.10926G>A	p.Met3642Ile	2
							58	g.336733G>T	c.9415G>T	p.Asp3139Tyr	4
26	**26**	M	Caucasian	0	Perinatal	−	11	g.15463A>G	c.764A>G	p.Tyr255Cys	2
27	**27**	F	Caucasian	0	Perinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							3	g.1733C>T	c.107C>T	p.Thr36Met	2
28	**28.1**	M	Caucasian	0	Perinatal	−	18	g.28168_28171delACTT	c.1626_1629delACTT	p.Leu542fs	1
							43	g.181333T>A	c.6992T>A	p.Ile2331Lys	2
28	**28.2**	F	Caucasian	13	Nonperinatal	−	18	g.28168_28171delACTT	c.1626_1629delACTT	p.Leu542fs	1
							43	g.181333T>A	c.6992T>A	p.Ile2331Lys	2
29	**29**	F	Caucasian	0	Perinatal	−	-	-	-	-	-
30	**30**	M	Caucasian	0	Perinatal	−	51	g.248470delG	c.8114delG	p.Gly2705fs	1
							13	g.19923T>C	c.920T>C	p.Ile307Thr	2
							55	g.312171A>G	c.8581A>G	p.Ser2861Gly	2
31	**31**	F	Caucasian	0.05	Perinatal	−	32	g.59994C>T	c.4870C>T	p.Arg1624Trp	2
							32	g.58871T>G	c.3747T>G	p.Cys1249Trp	2
32	**32**	M	Caucasian	0.1	Nonperinatal	+	61	g.425763delC	c.10955delC	p.Pro3652fs	1
							34	g.67422C>T	c.5498C>T	p.Ser1833Leu	4
33	**33.1**	F	Caucasian	0.1	Nonperinatal	−	38	g.172553T>G	c.6317T>G	p.Leu2106Arg	2
							12	g.18952A>G	c.874A>G	p.Ile292Val	4
							58	g.337106T>C	c.9788T>C	p.Val3219Ala	4
33	**33.2**	F	Caucasian	3	Nonperinatal	−	38	g.172553T>G	c.6317T>G	p.Leu2106Arg	2
							12	g.18952A>G	c.874A>G	p.Ile292Val	4
							58	g.337106T>C	c.9788T>C	p.Val3219Ala	4
34	**34**	F	Caucasian	0.2	Nonperinatal	−	16	g.26557C>A	c.1458C>A	p.Tyr486X	1
							30	g.56625A>G	c.3407A>G	p.Tyr1136Cys	2
							11	g.15436T>C	c.737T>C	p.Ile246Thr	3
35	**35**	F	Caucasian	0.2	Nonperinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
36	**36.1**	F	Caucasian	0.3	Nonperinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							58	g.336937G>A	c.9619G>A	p.Ala3207Thr	3

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

mily No	Patient No	Sex	Ethnic Origin	Age at diagnosis (y)	Presentation	CHF predominant	Exon	Genomic DNA	*PKHD1* variants Coding DNA	Protein	Pathogenicity score
36	**36.2**	F	Caucasian	0.4	Nonperinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							58	g.336937G>A	c.9619G>A	p.Ala3207Thr	3
37	**37**	M	Caucasian	0.3	Perinatal	−	40	g.175624_175625delTCinsCT	c.6655_6656delTCinsCT	p.Ser2219Leu	3
38	**38**	M	Caucasian	0.3	Perinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
39	**39**	F	Caucasian	0.3	Nonperinatal	−	32	g.58885_58886delCCinsG	c.3761_3762delCCinsG	p.Ala1254fs	1
							32	g.59994C>T	c.4870C>T	p.Arg1624Trp	2
40	**40**	M	Caucasian	0.4	Nonperinatal	−	16	g.26585C>T	c.1486C>T	p.Arg496X	1
							40	g.175639G>C	c.6670G>C	p.Gly2224Arg	2
41	**41**	F	African American	0.4	Nonperinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							58	g.337037G>A	c.9719G>A	p.Arg3240Gln	2
42	**42**	M	Caucasian	0.5	Nonperinatal	−	40	g.175598G>A	c.6629G>A	p.Gly2210Glu	2
							32	g.58871T>G	c.3747T>G	p.Cys1249Trp	2
43	**43.1***	F	Caucasian	0.7	Nonperinatal	−	60	g.340529delC	c.10136delC	pGly3378fs	1
							32	g.59994C>T	c.4870C>T	p.Arg1624Trp	2
43	**43.2***	F	Caucasian	28	Nonperinatal	−	60	g.340529delC	c.10136delC	pGly3378fs	1
							32	g.60258G>A	c.5134G>A	p.Gly1712Arg	2
44	**44**	F	Caucasian	0.8	Nonperinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							57	g.331653T>C	c.8870T>C	p.Ile2957Thr	2
45	**45**	F	Caucasian	0.8	Nonperinatal	−	58	g.336422delG	c.9104delG	p.Thr3035fs	1
							12	g.18956C>T	c.878C>T	p.Ala293Val	2
46	**46**	M	Caucasian	0.8	Nonperinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							9	g.13925A>G	c.664A>G	p.Ile222Val	2
47	**47**	F	Caucasian (French Canadian)		Nonperinatal	+	16	g.26496G>A	c.1397G>A	p.Gly466Glu	2
							61	g.425127T>A	c.10319T>A	p.Val3440Asp	4
48	**48**	M	Caucasian	1	Nonperinatal	+	-	-	-	-	-
49	**49**	F	Caucasian	1	Nonperinatal	+	-	-	-	-	-
50	**50**	M	Caucasian	1.1	Nonperinatal	−	7	g.11418G>A	c.474G>A	p.Trp158X	1
							32	g.59994C>T	c.4870C>T	p.Arg1624Trp	2
51	**51**	F	Caucasian	1.2	Nonperinatal	−	57	g.331653T>C	c.8870T>C	p.Ile2957Thr	2

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

mily No	Patient No	Sex	Ethnic Origin	Age at diagnosis (y)	Presentation	CHF predominant	Exon	Genomic DNA	*PKHD1* variants Coding DNA	Protein	Pathogenicity score
							55	g.312171A>G	c.8581A>G	p.Ser2861Gly	2
52	**52**	F	Caucasian	1.8	Nonperinatal	−	58	g.336937G>A	c.9619G>A	p.Ala3207Thr	3
53	**53**	F	Caucasian	2	Nonperinatal	+	6	g.8638A>G	c.428A>G	p.Tyr143Cys	3
54	**54**	F	Caucasian	2.7	Nonperinatal	−	46	g.201783A>T	c.7292A>T	p.Glu2431Val	2
55	**55**	M	Caucasian	3	Nonperinatal	−	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							16	g.26496G>A	c.1397G>A	p.Gly466Glu	2
56	**56.1**	M	Caucasian (French Canadian)	3	Nonperinatal	+	37	g.150800A>G	c.6097A>G	p.Arg2033Gly	2
							57	g.331653T>C	c.8870T>C	p.Ile2957Thr	2
56	**56.2**	F	Caucasian (French Canadian)	3	Nonperinatal	−	37	g.150800A>G	c.6097A>G	p.Arg2033Gly	2
							57	g.331653T>C	c.8870T>C	p.Ile2957Thr	2
57	**57**	F	Caucasian	3	Nonperinatal	+	22	g.34777G>A	c.2279G>A	p.Arg760His	2
							37	g.150728G>A	c.6025G>A	p.Ala2009Thr	2
58	**58**	M	African American	3	Nonperinatal	+	36	g.125051dupA	c.5895dupA	p.Leu1965fs	1
							32	g.60249C>T	c.5125C>T	p.Leu1709Phe	4
59	**59**	M	Caucasian	3.8	Nonperinatal	−	36	g.125051dupA	c.5895dupA	p.Leu1965fs	1
60	**60**	F	Caucasian	4	Nonperinatal	+	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							32	g.60249C>T	c.5125C>T	p.Leu1709Phe	4
61	**61**	M	Caucasian	5	Nonperinatal	−	16	g.26585C>T	c.1486C>T	p.Arg496X	1
							9	g.13925A>G	c.664A>G	p.Ile222Val	2
62	**62**	M	Caucasian	6	Nonperinatal	−	21	g.31775A>C	c.2057A>C	p.His686Pro	2
							22	g.34665C>T	c.2167C>T	p.Arg723Cys	4
63	**63**	F	Caucasian	6	Nonperinatal	+	3	g.1733C>T	c.107C>T	p.Thr36Met	2
							32	g.60345G>A	c.5221G>A	p.Val1741Met	2
64	**64**	F	Caucasian	6	Nonperinatal	−	57	g.331653T>C	c.8870T>C	p.Ile2957Thr	2
							35	g.74498T>G	c.5624T>G	p.Val1875Gly	3
65	**65**	F	African American	23	Nonperinatal	−	36	g.125051dupA	c.5895dupA	p.Leu1965fs	1
							18	g.28159T>C	c.1616T>C	p.Ile539Thr	2
66	**66.1**	M	Hispanic	28	Nonperinatal	+	57	g.331689_337692delTGTT	c.8906_8909delTGTT	p.Leu2969fs	1
							4	g.2535C>T	c.274C>T	p.Arg92Trp	3

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

mily No	Patient No	Sex	Ethnic Origin	Age at diagnosis (y)	Presentation	CHF predominant	Exon	Genomic DNA	*PKHD1* variants Coding DNA	Protein	Pathogenicity score
66	**66.2**	F	Hispanic	39	Nonperinatal	+	57	g.331689_337692delTGTT	c.8906_8909delTGTT	p.Leu2969fs	1
							4	g.2535C>T	c.274C>T	p.Arg92Trp	3
67	**67**	F	Caucasian	41	Nonperinatal	+	48	g.216882delC	c.7544delC	p.Ala2515fs	1
							34	g.67509C>T	c.5585C>T	p.Ser1862Leu	3
							65	g.452229G>T	c.11525G>T	p.Arg3842Leu	4
68	**68**	F	Caucasian	43	Nonperinatal	−	9	g.13925A>G	c.664A>G	p.Ile222Val	2
							65	g.452229G>T	c.11525G>T	p.Arg3842Leu	4

eeks gestation.

family is an aunt and niece pair.

hogenicity score key: 1 pathogenic (truncating); 2 probably pathogenic; 3 possibly pathogenic; 4 probably benign.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 4

PKHD1 sequence variants identified in the present study. Variants are classified as novel or previously reported and are presented with corresponding pathogenicity predictions. Reference sequences are genomic sequence NC000006.10 and mRNA sequence NM138694.

PKHD1 Sequence Variant				Number of Patient Alleles with the Variant	Segregation	Align GVGD Prediction 1	Align GVGD Prediction 2	PolyPhen Prediction	SNAP Prediction (accuracy)	Splice Variant Prediction	Frequency in Control Chromosomes	Overall Pathogenicity Score*
Exon	Genomic DNA	Coding DNA	Protein									
Novel Missense Variants												
4	g.2535C>T	c.274C>T	p.Arg92Trp	1	NA	Class 25	Class 25	Possibly pathogenic	Non-neutral (70%)	No change	1 in 212	3
11	g.15436T>C	c.737T>C	p.Ile246Thr	1	Consistent	Class 25	Class 65	Possibly pathogenic	Non-neutral (70%)	No change	0 in 240	3
11	g.15463A>G	c.764A>G	p.Tyr255Cys	1	Consistent	Class 65	Class 65	Probably pathogenic	Non-neutral (87%)	No change	0 in 400	2
12	g.18952A>G	c.874A>G	p.Ile292Val	1	Consistent	Class 0	Class 25	Benign	Neutral (78%)	No change	-	4
12	g.18956C>T	c.878C>T	p.Ala293Val	1	NA	Class 65	Class 65	Benign	Non-neutral (58%)	No change	0 in 400	2
16	g.26496G>A	c.1397G>A	p.Gly466Glu	4	Consistent	Class 65	Class 65	Probably pathogenic	Non-neutral (78%)	No change	0 in 400	2
16	g.26508G>T	c.1409G>T	p.Gly470Val	1	Consistent	Class 65	Class 65	Possibly pathogenic	Non-neutral (70%)	No change	0 in 400	2
18	g.28159T>C	c.1616T>C	p.Ile539Thr	1	NA	Class 65	Class 65	Possibly pathogenic	Non-neutral (70%)	No change	0 in 400	2
21	g.31775A>C	c.2057A>C	p.His686Pro	1	Consistent	Class 65	Class 65	Probably pathogenic	Non-neutral (70%)	No change	0 in 400	2
22	g.34665C>T	c.2167C>T	p.Arg723Cys	1	Inconsistent	Class 0	Class 0	Benign	Neutral (69%)	No change	-	4
22	g.34669C>G	c.2171C>G	p.Pro724Arg	1	NA	Class 0	Class 65	Probably pathogenic	Non-neutral (78%)	No change	0 in 400	3
32	g.60258G>A	c.5134G>A	p.Gly1712Arg	1	NA	Class 65	Class 65	Probably pathogenic	Non-neutral (87%)	No change	0 in 400	2
34	g.67374T>G	c.5450T>G	p.Val1817Gly	2	NA	Class 35	Class 35	Possibly pathogenic	Non-neutral (63 %)	No change	0 in 202	2
35	g.74498T>G	c.5624T>G	p.Val1875Gly	1	NA	Class 65	Class 65	Possibly pathogenic	Non-neutral (78%)	No change	1 in 400	3
36	g.124939G>T	c.5783G>T	p.Trp1928Leu	1	NA	Class 55	Class 55	Probably pathogenic	Non-neutral (93%)	No change	0 in 400	2
37	g.150728G>A	c.6025G>A	p.Ala2009Thr	1	NA	Class 55	Class 55	Benign	Non-neutral (82%)	No change	0 in 400	2
38	g.172553T>G	c.6317T>G	p.Leu2106Arg	2	Consistent	Class 65	Class 65	Possibly pathogenic	Non-neutral (93%)	No change	0 in 400	2
40	g.175598G>A	c.6629G>A	p.Gly2210Glu	1	Consistent	Class 65	Class 65	Probably pathogenic	Non-neutral (96%)	No change	0 in 400	2
40	g.175624_175625delTCinsCT	c.6655_6656delTCinsCT	p.Ser2219Leu	1	Consistent	Class 15	Class 65	Benign	Non-neutral (58%)	No change	0 in 222	3
40	g.175639G>C	c.6670G>C	p.Gly2224Arg	1	NA	Class 65	Class 65	Probably pathogenic	Non-neutral (96%)	May change splicing	0 in 400	2
46	g.201755T>C	c.7264T>C	p.Cys2422Arg	1	Consistent	Class 65	Class 65	Probably pathogenic	Non-neutral (82%)	No change	0 in 400	2
46	g.201783A>T	c.7292A>T	p.Glu2431Val	1	Consistent	Class 65	Class 65	Possibly pathogenic	Non-neutral (58%)	May change splicing	0 in 400	2

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

PKHD1 Sequence Variant				Number of Patient Alleles with the Variant	Segregation	Align GVGD Prediction 1	Align GVGD Prediction 2	PolyPhen Prediction	SNAP Prediction (accuracy)	Splice Variant Prediction	Frequency in Control Chromosomes	Overall Pathogenicity Score*
Exon	**Genomic DNA**	**Coding DNA**	**Protein**									
48	g.217055C>T	c.7717C>T	p.Arg2573Cys	1	NA	Class 45	Class 65	Probably pathogenic	Non-neutral (87%)	No change	0 in 400	2
50	g.237033T>C	c.7981T>C	p.Tyr2661His	1	Inconsistent	Class 0	Class 65	Possibly pathogenic	Neutral (53%)	No change	-	4
53	g.293651T>G	c.8393T>G	p.Val2798Gly	1	NA	Class 65	Class 65	Possibly pathogenic	Non-neutral (70%)	No change	0 in 400	2
53	g.293665T>C	c.8407T>C	p.Cys2803Arg	1	NA	Class 65	Class 65	Probably pathogenic	Non-neutral (82%)	No change	0 in 400	2
53	g.293669T>A	c.8410T>A	p.Met2804Lys	1	NA	Class 55	Class 55	Probably pathogenic	Non-neutral (82%)	No change	0 in 400	2
58	g.336464A>G	c.9146A>G	p.His3049Arg	2	Consistent	Class 0	Class 25	Probably pathogenic	Non-neutral (58%)	No change	0 in 400	3
58	g.336937G>A	c.9619G>A	p.Ala3207Thr	2	Consistent	Class 0	Class 55	Benign	Non-neutral (58%)	No change	0 in 400	3
58	g.337106T>C	c.9788T>C	p.Val3219Ala	1	Consistent	Class 0	Class 0	Benign	Non-neutral (78%)	No change	-	4
67	g.465497C>T	c.11869C>T	p.Arg3957Cys	1	Consistent	Class 0	Class 65	Probably pathogenic	Non-neutral (82%)	No change	0 in 214	3
Previously Reported Missense Variants												
3	g.1733C>T	c.107C>T	p.Thr36Met	21	Consistent	Class 65	Class 65	Possibly pathogenic	Non-neutral (87%)	-	-	2
6	g.8638A>G	c.428A>G	p.Tyr143Cys	1	NA	Class 15	Class 65	Possibly pathogenic	Non-neutral (78%)	-	-	3
9	g.13925A>G	c.664A>G	p.Ile222Val	6	Consistent	Class 25	Class 65	Benign	Non-neutral (58%)	-	-	2#
13	g.19923T>C	c.920T>C	p.Ile307Thr	1	Consistent	Class 65	Class 25	Possibly pathogenic	Non-neutral (87 %)	-	-	2
22	g.34777G>A	c.2279G>A	p.Arg760His	1	NA	Class 0	Class 0	Benign	Non-neutral (82%)	-	-	2#
30	g.56625A>G	c.3407A>G	p.Tyr1136Cys	1	Consistent	Class 65	Class 0	Probably pathogenic	Non-neutral (82%)	-	-	2
30	g.56685C>T	c.3467C>T	p.Ser1156Leu	1	Consistent	Class 0	Class 65	Benign	Non-neutral (70%)	-	-	2
32	g.58871T>G	c.3747T>G	p.Cys1249Trp	2	Consistent	Class 65	Class 25	Probably pathogenic	Non-neutral (96%)	-	-	2
32	g.59994C>T	c.4870C>T	p.Arg1624Trp	4	Consistent	Class 25	Class 65	Benign	Neutral (69%)	-	-	2#
32	g.60249C>T	c.5125C>T	p.Leu1709Phe	2	Consistent	Class 15	Class 25	Benign	Non-neutral (58%)	-	-	4#
32	g.60345G>A	c.5221G>A	p.Val1741Met	1	NA	Class 15	Class 15	Benign	Non-neutral (70%)	-	-	2#
34	g.67422C>T	c.5498C>T	p.Ser1833Leu	1	NA	Class 15	Class 15	Possibly pathogenic	Non-neutral (82%)	-	-	4
34	g.67509C>T	c.5585C>T	p.Ser1862Leu	1	NA	Class 0	Class 65	Possibly pathogenic	Non-neutral (78%)	-	-	3
37	g.150800A>G	c.6097A>G	p.Arg2033Gly	1	Consistent	Class 65	Class 65	Probably pathogenic	Non-neutral (87%)	-	-	2
43	g.181333T>A	c.6992T>A	p.Ile2331Lys	2	Consistent	Class 35	Class 65	Possibly pathogenic	Non-neutral (82%)	-	-	2

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

PKHD1 Sequence Variant				Number of Patient Alleles with the Variant	Segregation	Align GVGD Prediction 1	Align GVGD Prediction 2	PolyPhen Prediction	SNAP Prediction (accuracy)	Splice Variant Prediction	Frequency in Control Chromosomes	Overall Pathogenicity Score*
Exon	Genomic DNA	Coding DNA	Protein									
50	g.237120T>C	c.8068T>C	p.Trp2690Arg	1	NA	Class 65	Class 55	Probably pathogenic	Non-neutral (87%)	-	-	2
52	g.254017_254018delGGinsCC	c.8246_8247delGGinsCC	p.Trp2749Ser	1	NA	Class 65	Class 55	Probably pathogenic	Non-neutral (82%)	-	-	2
55	g.312171A>G	c.8581A>G	p.Ser2861Gly	3	NA	Class 0	Class 65	Benign	Neutral (69%)	-	-	2#
57	g.331653T>C	c.8870T>C	p.Ile2957Thr	5	Consistent	Class 0	Class 65	Possibly pathogenic	Non-neutral (78%)	-	-	2#
58	g.336733G>T	c.9415G>T	p.Asp3139Tyr	1	NA	Class 15	Class 65	Benign	Non-neutral (70%)	-	-	4#
58	g.337037G>A	c.9719G>A	p.Arg3240Gln	1	NA	Class 0	Class 35	Benign	Neutral (53%)	-	-	2#
61	g.425127T>A	c.10319T>A	p.Val3440Asp	1	NA	Class 0	Class 35	Benign	Non-neutral (82%)	-	-	4#
61	g.425734G>A	c.10926G>A	p.Met3642Ile	1	NA	Class 0	Class 15	Benign	Neutral (78%)	-	-	2#
65	g.452229G>T	c.11525G>T	p.Arg3842Leu	2	NA	Class 0	Class 0	Probably pathogenic	Non-neutral (87%)	-	-	4#
Novel Truncating Variants												
16	g.26557C>A	c.1458C>A	p.Tyr486X	1	Consistent	-	-	-	-	-	-	1
18	g.28168_28171delACTT	c.1626_1629delACTT	p.Leu542fs	1	Consistent	-	-	-	-	-	-	1
32	g.58890delC	c.3766delC	p.Gln1256fs	1	NA	-	-	-	-	-	-	1
45	g.198973delT	c.7120delT	p.Phe2374fs	1	NA	-	-	-	-	-	-	1
48	g.216882delC	c.7544delC	p.Ala2515fs	1	NA	-	-	-	-	-	-	1
57	g.331689_337692delTGTT	c.8906_8909delTGTT	p.Leu2969fs	1	NA	-	-	-	-	-	-	1
58	g.336422delG	c.9104delG	pThr3035fs	1	NA	-	-	-	-	-	-	1
60	g.340529delC	c.10136delC	pGly3378fs	1	NA	-	-	-	-	-	-	1
61	g.425763delC	c.10955delC	p.Pro3652fs	1	NA	-	-	-	-	-	-	1
Previously Reported Truncating Variants												
7	g.11418G>A	c.474G>A	p.Trp158X	1	NA	-	-	-	-	-	-	1
16	g.26585C>T	c.1486C>T	p.Arg496X	2	NA	-	-	-	-	-	-	1
23	g.36376C>T	c.2341C>T	p.Arg781X	1	Consistent	-	-	-	-	-	-	1
32	g.58885_58886delCCinsG	c.3761_3762delCCinsG	p.Ala1254fs	1	Consistent	-	-	-	-	-	-	1
36	g.125051dupA	c.5895dupA	p.Leu1965fs	3	Consistent	-	-	-	-	-	-	1

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

PKHD1 Sequence Variant				Number of Patient Alleles with the Variant	Segregation	Align GVGD Prediction 1	Align GVGD Prediction 2	PolyPhen Prediction	SNAP Prediction (accuracy)	Splice Variant Prediction	Frequency in Control Chromosomes	Overall Pathogenicity Score*
Exon	Genomic DNA	Coding DNA	Protein									
51	g.248470delG	c.8114delG	p.Gly2705fs	2	NA	-	-	-	-	-	-	1
58	g.337007delA	c.9689delA	p.Asp3230fs	1	Consistent	-	-	-	-	-	-	1
61	g.425261dupT	c.10452dupT	p.Phe3485fs	1	Consistent	-	-	-	-	-	-	1
61	g.425436_425444del8	c.10628_10635del8	p.Leu3543fs	1	Consistent	-	-	-	-	-	-	1
61	g.425445delT	c.10637delT	p.Val3546fs	1	NA	-	-	-	-	-	-	1
Previously Reported Splice Variants												
IVS39	-	c.6490+2T>C	-	1	NA	-	-	-	-	Changes splicing		2
IVS55	-	c.8642+1G>A**	-	1	Consistent	-	-	-	-	Changes splicing	-	2
Novel In-Frame Deletion												
58	g.336369_336371delATC	c.9048_9050delATC	p.Ser3017del	1	Consistent	-	-	-	-	-	-	2

* Overall pathogenicity score key: 1 pathogenic (truncating); 2 probably pathogenic; 3 possibly pathogenic; 4 probably benign.

** This splice mutation was previously reported in the same patient.

Overall pathogenicity scores for these variants are based on the published data.