Milder Presentation of Recessive Polycystic Kidney Disease Requires Presence of Amino Acid Substitution Mutations

LASZLO FURU*, LUIZ F. ONUCHIC,[‡] ALI GHARAVI,[†§] XIAOYING HOU,[‖] ERNIE L. ESQUIVEL,* YASUYUKI NAGASAWA,[¶] CARSTEN BERGMANN,[#] JAN SENDEREK,[#] ELLIS AVNER,** KLAUS ZERRES,[#] GREGORY G. GERMINO,[¶] LISA M. GUAY-WOODFORD[‖], and STEFAN SOMLO*[†]

*Departments of *Internal Medicine and [†]Genetics, Yale University School of Medicine, New Haven, Connecticut; [‡]Department of Medicine, University of Sao Paulo, Sao Paulo, Brazil; [§]Department of Internal Medicine, Mt. Sinai School of Medicine, New York, New York; [‖]Departments of Medicine and Pediatrics, University of Alabama at Birmingham, Birmingham, Alabama; [¶]Departments of Medicine and Genetics, Johns Hopkins University, Baltimore, Maryland; #Institute of Human Genetics, Technical University of Aachen, Aachen, Germany; and **Department of Pediatrics, Rainbow Babies Childrens Hospital, Cleveland, Ohio.*

Abstract. Autosomal recessive polycystic kidney disease (ARPKD; MIM 263200) is a hereditary and severe form of polycystic disease affecting the kidneys and biliary tract with an estimated incidence of 1 in 20,000 live births. The clinical spectrum is widely variable: up to 50% of affected neonates die shortly after birth, whereas others survive to adulthood. Mutations at a single locus, polycystic kidney and hepatic disease 1 (*PKHD1*), are responsible for all typical forms of ARPKD. Mutation detection was performed in *PKHD1* by DHPLC in 85 affected, unrelated individuals. Seventy-four amplicons were amplified and analyzed from the *PKHD1* genomic locus. Sequence variants were considered pathogenic when they were not observed in 160 control individuals (320 chromosomes). For purposes of genotype-phenotype comparisons, families were stratified by clinical presentation into two groups: the severe perinatal group, in which at least one affected child presented with perinatal disease and neonatal demise, and the less severe, nonperinatal group, in which none of the affected children died in the neonatal period. Forty-one mutations were found in 55 affected disease chromosomes; 32 of these mutations have not been reported previously. Mutations were distributed throughout the portions of gene encoding the predicted extracellular portion of the protein product. The most commonly encountered mutation, T36M, was found in 8 of 55 disease chromosomes. Amino acid substitutions were found to be more commonly associated with a nonlethal presentation, whereas chain terminating mutations were more commonly associated with neonatal demise ($\chi^2 = 11.54$, $P = 0.003$). All patients who survive the neonatal period have at least one amino acid substitution mutation, suggesting that such substitutions produce milder disease through production of partially functional protein products. The nature of the germline mutations in ARPKD plays a significant role in determining clinical outcome.

Autosomal recessive polycystic kidney disease (ARPKD; MIM 263200) is a hereditary and severe form of polycystic disease affecting the kidneys and biliary tract with an estimated incidence of 1 in 20,000 live births (1). The clinical spectrum is widely variable: up to 50% of affected neonates die shortly after birth, whereas others survive to adulthood (2,3). Mutations at a single locus, polycystic kidney and hepatic disease 1 (*PKHD1*), are responsible for all typical forms of ARPKD (4,5). It has been proposed that the relative concordance of phenotypes within the majority of families and the wide variation between families (6) are the result of allelic heterogeneity at the *PKHD1* locus (1,7). The discordance of phenotypes within a minority of sibships (8,9) may represent the additional modifying effects of other genetic loci and environmental factors (2).

The clinical spectrum of ARPKD is widely variable, with most cases presenting in infancy. Children with early and severe presentations typically exhibit markedly enlarged echogenic kidneys and oligohydramnios in association with intrauterine renal failure. As a result of the oligohydramnios, approximately 50% of affected neonates have the Potter sequence, with pulmonary hypoplasia, a characteristic facies, and deformities of the spine and limbs. At birth, these neonates often have a critical degree of pulmonary hypoplasia and die within the first few hours of life as a result of respiratory insufficiency. In affected neonates, the kidneys are symmetrically enlarged and cysts appear as fusiform dilations of col-

Received February 25, 2003. Accepted April 24, 2003.

Correspondence to Dr. Stefan Somlo, Yale University School of Medicine, P.O. Box 208029, 300 Cedar Street, New Haven, CT 06520-8029. Phone: 203-737-2974; Fax: 203-785-4904; E-mail: stefan.somlo@yale.edu and Dr. Lisa M. Guay-Woodford, University of Alabama at Birmingham, Kaul Building 740, 720 South 20th Street, Birmingham, AL 35294-0024. Phone: 205-934-7308; Fax: 205-975-5689; E-mail: lgw@uab.edu

A.G. is currently affiliated with Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, New York.

1046-6673/1408-2004

Journal of the American Society of Nephrology

DOI: 10.1097/01.ASN.0000078805.87038.05

lecting ducts extending radially from the renal pelvis to the cortex (2). The ectatic collecting ducts are lined by cuboidal epithelia that retain lectin-binding properties of the segments of origin (10). In less severely affected children, the extent and degree of collecting duct dilation is more limited and the course is more chronic with accompanying loss of glomeruli, tubular atrophy, and interstitial fibrosis (2). A minority of affected individuals survive well into adulthood (3). The principal causes of morbidity and mortality in affected children are hypertension, progressive renal insufficiency, and portal tract fibrosis (2). Hypertension, which can be severe, occurs in 75 to 100% of patients and can occur in the setting of preserved renal function. Among children who survive the immediate perinatal period, approximately 50% progress to ESRD in the first decade (11). Congenital hepatic fibrosis results from ductal plate malformation that leads to biliary dysgenesis (12). Portal tract fibrosis and portal hypertension are the significant clinical sequelae. There is substantial variation in the relative severity of presentation of the renal and hepatic components of the ARPKD phenotype (2,8). However, all patients have histopathologic presence of congenital hepatic fibrosis, and some physicians require this finding to make a diagnosis of ARPKD.

PKHD1 was recently discovered as a novel gene that extends over approximately 500 kb on chromosome 6p12 and is strongly expressed in fetal and adult kidney, with weaker expression in liver and other tissues (13,14). The longest transcript contains 67 exons with an open reading frame composed of 66 exons encoding a 4074–amino acid protein, polyductin/fibrocystin (13,14). It is predicted to have a single transmembrane-spanning domain very near its carboxyl terminus and multiple IPT/TIG domains and PbH1 repeats in its predicted approximately 3860–amino acid extracellular NH_2-terminus (13,14). ARPKD results from partial or complete loss of polyductin/fibrocystin function (13,14). The structure and domain composition of the deduced protein combined with knowledge of the human disease phenotype suggests that polyductin/fibrocystin is a cell-signaling molecule (receptor, co-receptor, or ligand) that is essential for maintaining normal development and function of renal tubules and bile ducts. The presence of the PbH1 domains shared with proteins of the pectin lyase superfamily suggests that polyductin/fibrocystin may have catalytic activity on polysaccharide substrates.

The advent of the discovery of *PKHD1* has permitted the definition of causative mutations in affected families and the potential for correlation of these mutations with the diverse spectrum of interfamilial clinical variation. In the current study of an ethnically diverse population, we describe 41 mutations in 55 affected disease chromosomes; 32 of these mutations have not been reported previously (13–15). Mutations were distributed throughout the region of gene encoding the predicted extracellular portion of the protein product. We found that amino acid substitutions were more commonly associated with a nonlethal presentation, whereas chain-terminating mutations were more commonly associated with neonatal demise. All patients who survived the neonatal period had at least one amino acid substitution mutation, suggesting that some amino acid substitution mutations produce milder disease and may be associated with a better prognosis.

Materials and Methods

Patients and Samples

Families and individuals were ascertained from databases at the University of Alabama at Birmingham and the Rheinishch-Westfalische Technishce Hoechschule (Aachen, Germany). ARPKD was diagnosed according to the criteria of Zerres *et al.* (1). Study subjects and family members provided informed consent by approved protocols at the recruiting institutions. Study subjects represented the full spectrum of clinical presentations for ARPKD (Table 1). Genomic DNA was prepared from whole blood by standard protocols (13).

Mutation Detection

We PCR-amplified 74 amplicons from the *PKHD1* genomic locus using a set of primers designed to amplify exon sequences, the adjacent splice sites, and approximately 40 nucleotides of flanking sequence on each side (15). Several primers were modified for the current study (Table 2). Amplified fragment sizes ranged from 149 to 587 nucleotide bases. Exons 32, 59, and 65 were amplified in 4, 3, and 3 overlapping PCR products, respectively, to limit the size of the amplicons. PCR product sequences were optimized for DHPLC by selection of primer pairs that minimized the temperature difference in melting domains across the region as determined by the Wavemaker version 4.1 system control software (Transgenomic, Inc., Omaha, NE). Amplifications were performed using 96-well plates in a Perkin Elmer 9700 thermal cycler with the following conditions: initial denaturation at 94°C for 4 min followed by 35 cycles of 30 s each at 94°C, at the amplicon-specific annealing temperature (Table 2) (15), and at 72°C, followed by a final extension at 72°C for 7 min. Standard PCR reactions were prepared in 30 μl using 20 ng of genomic DNA and 0.5 unit of either regular *Taq* or AmpliTaq DNA polymerase (PE Biosystems, Foster City, CA). The reaction volume provided amplification product for two rounds of DHPLC analysis as well as for sequencing templates.

Mutation detection was performed by heteroduplex analysis using the Transgenomic Wave DHPLC system (Transgenomic, Inc.). PCR products were denatured at 98°C for 4 min and allowed to reanneal in the thermocycler block for 30 min. Samples were placed into the DHPLC autosampler and held at 10°C as 8 to 12 μl of each amplicon was injected onto the column. The mobile phase consisted of a mixture of Buffers A (0.1 M TEAA and 1 mM EDTA) and B (25% acetonitrile in 0.1 M TEAA). The buffer gradient and optimum denaturing temperature required for successful resolution of heteroduplexes for each amplicon were determined according to the Wavemaker version 4.1 system control software. Samples were eluted with a linear acetonitrile gradient at a flow rate of 0.9 ml/min over 7 min and routinely analyzed at two denaturing temperatures. Samples showing altered elution properties not present in controls were purified (Qiaex; Qiagen, Inc., Valencia, CA) and sequenced in both directions with BigDye Deoxy Terminator cycle sequencing on an ABI 3700 DNA sequencer (Applied Biosystems, Foster City, CA). The PCR primers were also used for sequencing reactions. Sequence variations were identified by visual inspection and comparison of the resulting electropherograms (Editview, version 1.1; Applied Biosystems). When amplicons had altered elution properties in both control and patient DNA, both samples were sequenced to confirm their identity at the nucleotide level. All putative mutations were tested by segregation analysis when family material was available.

Table 1. Newly identified mutations in *PKHD1*

ID	Exon[a]	Nucleotide Change[b]	ORF Change[c]	Age[d] (at Diagnosis or Death)	Sys. HTN	Kidney Disease	Portal HTN	Ethnic Origin
AL 7	3	107C→ T	T36M	Perinatal (neonatal death)	No	Nephromegaly	No	American-Caucasian
AL 13	3	107C→ T	T36M	Perinatal (neonatal death)	No	Nephromegaly	No	American-Caucasian
AL 24	3	107C→ T	T36M	Prenatal diagnosis with termination	No	Nephromegaly	No	Unknown
	32	3761_3762delCCinsG	A1254Xfs					
167/704	3	107C→ T	T36M	Perinatal (neonatal death)	No	Nephromegaly	No	German/Romanian
338/1389	3	107C→ T	T36M	Perinatal (neonatal death)	No	Nephromegaly	No	English/Dutch
313/1289	3	107C→ T	T36M	Perinatal (neonatal death); sibling w/infantile	Yes	CRI in second decade	Yes; in 20-yr-old sibling	German
	33	5365G→ C[e]	V1789L					
AL 19	IVS5	IVS5 + 1G→ T[e]	splice site	Perinatal (neonatal death)	No	Nephromegaly	No	American-Caucasian
AL 40	14	977G→ T[e]	G326V	Perinatal (neonatal death)	No	Nephromegaly	No	American-Caucasian
AL 8	15	1159_1161delAAT[e]	N387del	Perinatal (neonatal death)	No	Nephromegaly	No	American-Caucasian
	36	5825A→ G[e]	D1942G					
AL 23	26	2810G→ A[e]	W937X	Perinatal (neonatal death)	No	Nephromegaly	No	Asian/Caucasian
	61 (65)	11074C→ T[e]	R3692X					American
388/1597	26	2810G→ A[e]	W937X	Perinatal (neonatal death)	No	Nephromegaly	Unknown	German
AL 14	29	3306delT	Y1102X	Perinatal (neonatal death)	No	Nephromegaly	No	American-Caucasian
AL 15	30	3367G→ A	G1123S	Perinatal (neonatal death)	No	Nephromegaly	No	American-Caucasian
	32	4751G→ T[e]	S1584I					
AL 17	32	3761_3762delCCinsG	A1254Xfs	Perinatal (neonatal death)	No	Nephromegaly	No	American-Caucasian
AL 21	32	3761_3762delCCinsG	A1254Xfs	Perinatal (neonatal death)	No	Nephromegaly	No	American-Caucasian
	50 (51)	7921A→ G[e]	T2641A					
95/1862	32	3762_3763insG[e]	P1255Xfs	Perinatal (neonatal death)	Yes	Nephromegaly	Unknown	Turkish
AL 39	IVS39	IVS39 + 2T→ G[e]	splice site	Perinatal	Yes	ESRF in second decade	No	American-Caucasian
	(IVS40)			Sibling w/perinatal (neonatal death)				
367/1515	46 (47)	7264T→ G[e]	C2422G	Perinatal (neonatal death)	No	Nephromegaly	No	Australian
398/1639	47 (48)	7477C→ T[e]	Q2493X	Perinatal (neonatal death)	No	Nephromegaly	Unknown	Dutch
	58 (59)	9499A→ T[e]	I3167L					
AL 34	57 (58)	8870T→ C	I2957T	Prenatal diagnosis with termination	No	Nephromegaly	No	American-Caucasian
	61 (65)	10637_10638insC[e]	L3547Xfs					
AL 50	57 (58)	8829_8830insC	I2944Xfs	Perinatal (neonatal death)	No	Nephromegaly	No	Scottish-Caucasian
AL 25	58 (59)	9530T→ C[e]	I3177T	Prenatal diagnosis with termination	No	Nephromegaly	No	Unknown
296/1233	58 (59)	9370C→ T[e]	H3124Y	Perinatal (neonatal death)	No	Nephromegaly	No	Italian
373/1544	59 (60)	9878A→ T	D3293V	Perinatal (neonatal death)	No	Nephromegaly	No	Turkish
AL 37	60 (61)	10075delG	G3359Xfs	Prenatal diagnosis with termination; sibling w/perinatal (neonatal death)	No	Nephromegaly	No	American-Caucasian
	61 (65)	10728G→ A[e]	W3576X					
AL 30	61 (65)	10765C→ T[e]	Q3589X	Perinatal (neonatal death)	No	Nephromegaly	No	Hispanic-American
163/699	3	107C→ T	T36M	Infantile	Yes	Nephromegaly	Yes; Caroli disease	German
371/1535	3	107C→ T	T36M	Childhood	No	Nephromegaly	Yes	Austrian
	46 (47)	7264T→ G[e]	C2422G					
AL 35	25	2695A→ C[e]	T899P	Infantile	No	Nephromegaly	Unknown	Turkish

Table 1. Continued

ID	Exon[a]	Nucleotide Change[b]	ORF Change[c]	Age[d] (at Diagnosis or Death)	Sys. HTN	Kidney Disease	Portal HTN	Ethnic Origin
	57 (58)	8829_8830insC	I2944Xfs					
AL 4	32	5125C→ T[e]	L1709F	Infantile	Yes	Nephromegaly; CRI in second decade	Yes	South African-Afrikaner
	55 (56)	8588A→ G[e]	Y2863G					
AL 58	33	5342C→ T[e]	T1781I	Infantile	Yes	Nephromegaly; CRI	Yes	American-Caucasian
AL 3	37	5912G→ A[e]	G1971D	Infantile	Yes	Nephromegaly; ESRF in mid-20s w/transplant	No; +hepatomegaly	American-Caucasian
	55 (56)	8588A→ G[e]	Y2863G					
AL 27	37	6094G→ C[e]	V2032L	Infantile	Yes	Nephromegaly; ESRF in second decade w/ transplant	No; +hepatomegaly	American-Caucasian
AL 46	50 (51)	8011C→ T	R2671X	Infantile	Yes	Nephromegaly	Unknown	Saudi Arabian
	57 (58)	8948C→ T[e]	S2983L					
AL 41	53 (54)	8315T→ C[e]	L2772P	Infantile	Yes	Nephromegaly	Unknown	Turkish
AL 33	55 (56)	8581A→ G[e]	S2861G	Childhood	No	None	Yes	American-Caucasian
389/1600	57 (58)	8829_8830insC	I2944Xfs	Childhood	Yes	CRI, second decade	Unknown	Czech
AL 31	58 (59)	9107T→ G[e]	V3036G	Childhood	No	None	Yes	Hispanic-American
AE 241	58 (59)	9524A→ G[e]	N3175S	Childhood	Yes	Nephromegaly	No	African American/Caribbean
331/1360	58 (59)	9523A→ G[e]	N3175D	Infantile	—	Unknown	Unknown	Turkish
AL 53	63 (67)	11347C→ T[e]	P3783S	Infantile	Yes	Nephromegaly	Unknown	Saudi Arabian

Sys. HTN, systemic hypertension; portal HTN, portal hypertension; CRI, chronic renal insufficiency; ESRF, end-stage renal failure requiring renal replacement therapy.

[a] Exons are numbered sequentially from 1–67 of the longest transcript containing the putative ORF; where these exon numbers differ from the total of 71 non-overlapping exons described in our original report, Onuchic et. al., we provide the corresponding alternative exon number in parentheses.

[b] Nucleotides are numbered with the "A" in the initiation ATG as 1 and are based on 67 exon transcript that contains longest ORF.

[c] Codons are numbered based on the 4074 amino acid sequence of the longest ORF.

[d] Clinical findings: All patients must have congenital hepatic fibrosis but not all were specifically diagnosed by histopathology; those with known portal hypertension are indicated. Among patients with perinatal presentation and neonatal demise, all have nephromegaly and most had rapid and severe course during which systemic hypertension may not have established itself; hence, absence of Sys HTN in this group needs to be interpreted in light of the clinical course.

[e] Mutations described for the first time in the current report.

Table 2. Primers for mutation detection in *PKHD1*[a]

Exon	Forward	Reverse	Size (bp)	T_m (°C)[b]
27	5′-TCACTGAGAGATTATTTTATTTGTTTTT-3	5-TCAGAGGGTCAGACATACTGTGA-3	365	58, 60
32A	5-CACATGCCCTACCTTCCACT-3	5-CTCCCACATGCAGGCTCA-3	442	58, 60
32E	5-TATGCCCCAAGTGTTTCATTATTT-3	5-GTGAAAGGAGCTACCAATTCATTT-3	587	57, 58
33	5-GTAAAGAAGGGAGATTTGCCTGTA-3	5-GAATTAACCAAAGAATATCATTTCCA-3	267	57, 59
34	5-GGCCTTAGGTTCTCTCTGTGG-3	5-AAGTTCAGGGAGGGAGAAGG-3	301	59, 61
39	5′-TTTGGAGTGATGTCCTCAGTTCT-3′	5′-GCAATGCCATCTATCATCAGAC-3′	279	59, 61
60	5-TTCGATATTTGTGGCTGGTG-3	5-GGGTTTGAAGAATTGCCAAG-3	297	53, 55

[a] Redesigned from (13).
[b] DHPLC column temperature, two conditions.

Statistical Analyses

Genotype-phenotype comparisons were performed using χ^2 analysis with continuity correction; Fisher exact tests were computed when sample size was small. In our primary analysis, we compared allelic and genotypic frequencies of chain-terminating and amino acid substitution mutations among our patients (current report and (13)) with and without perinatal presentation. We extended this analysis to include patients reported in the literature (14,15). With four primary evaluations, $P < 0.0125$ was considered significant after Bonferroni correction for multiple comparisons. Statistical analyses were performed using SPSS 11.0.

Results

Mutation Detection in PKHD1

DHPLC is well suited for discovery of heterozygous mutations but may be less robust for detection of homozygous mutations in the absence of mixing strategies. However, because most ARPKD patients seem to be compound heterozygotes in the absence of known consanguinity or founder mutations (13–15), we expected DHPLC to be a good mutation detection modality for this study even without mixing studies. We performed mutation detection in *PKHD1* by DHPLC in 85 affected, unrelated individuals. Among these, 60 samples underwent complete screening of the 67 exons that compose the transcript with the longest predicted ORF of *PKHD1*. Another 25 individual samples underwent partial screening of an average of 42 exons of the gene as a result of insufficient genomic DNA template to complete the screen. In total, we identified 49 pathogenic sequence variants affecting 75 disease chromosomes (current study and (13)). The criteria for calling a sequence variant a mutation included a predicted chain-termination effect on the translation product, segregation with the disease phenotype, and absence of the variant in 160 control individuals (320 chromosomes). Among the 60 samples that were screened completely, mutations were found on both chromosomes in 22 individuals (37%), mutation on one chromosome was found in 18 samples (30%), and no mutations were found in 20 samples (33%). Overall, when the entire gene was screened, mutations were found in 62 of 120 disease chromosomes, approximately 52% detection efficiency. Eight of the patients in whom both mutations were found and four with one mutation have been reported previously (13). Not surprising, when less than the entire gene was screened, the detection was less efficient: 13 (52%) with one mutation found and 12 (48%) with no mutations found.

We found 41 different mutations in 55 disease chromosomes to go along with the 20 mutations from the same study that we reported previously (Table 1; Figure 1 (13)). Among the 55 disease chromosomes, 36 chromosomes contained 32 novel mutations that have not been described previously (13–15). Our study population was ethnically diverse (Table 1 and (13)). We did not find evidence for founder mutations in this population, with the possible exception of the previously reported R1624W amino acid substitution seen in Saudi Arabian patients (13). Several mutations were identified in multiple unrelated individuals (Figure 1). Most notable, the previously described T36M variant stood out as a particularly common mutation in our population. It occurred in 8 of the 55 disease chromosomes described in this report and in 9 of the 75 chromosomes when we combined it with our previous report (13). T36M was found in seemingly unrelated individuals across national boundaries and has been described in each study reporting *PKHD1* mutations to date (13–15). Others have suggested that it has recurred *de novo* a number of times (15) and may represent a mutational "hot spot." We cannot rule out a founder effect for T36M as it seems to be particularly highly represented in German and central European populations in both this study and previous reports (15). Given the potential limitation of DHPLC in detecting homozygous variants and the finding that 10 to 15% of our heterozygous mutant chromosomes had T36M, we considered the possibility that some patients in whom no mutations were found were in fact homozygous for T36M. We excluded homozygous T36M mutations by direct sequencing in the 32 samples in which no mutations were discovered.

Mutations are distributed throughout the gene but confined to the predicted extracellular portion of the predicted protein. There is a preponderance of amino acid substitution changes. We found 46 (61%) amino acid substitution changes and 29 (39%) predicted chain-terminating mutations among mutations detected in 75 disease chromosomes. All amino acid substitutions described as mutations were absent in 320 control chromosomes. Of the 32 different amino acid substitutions that we

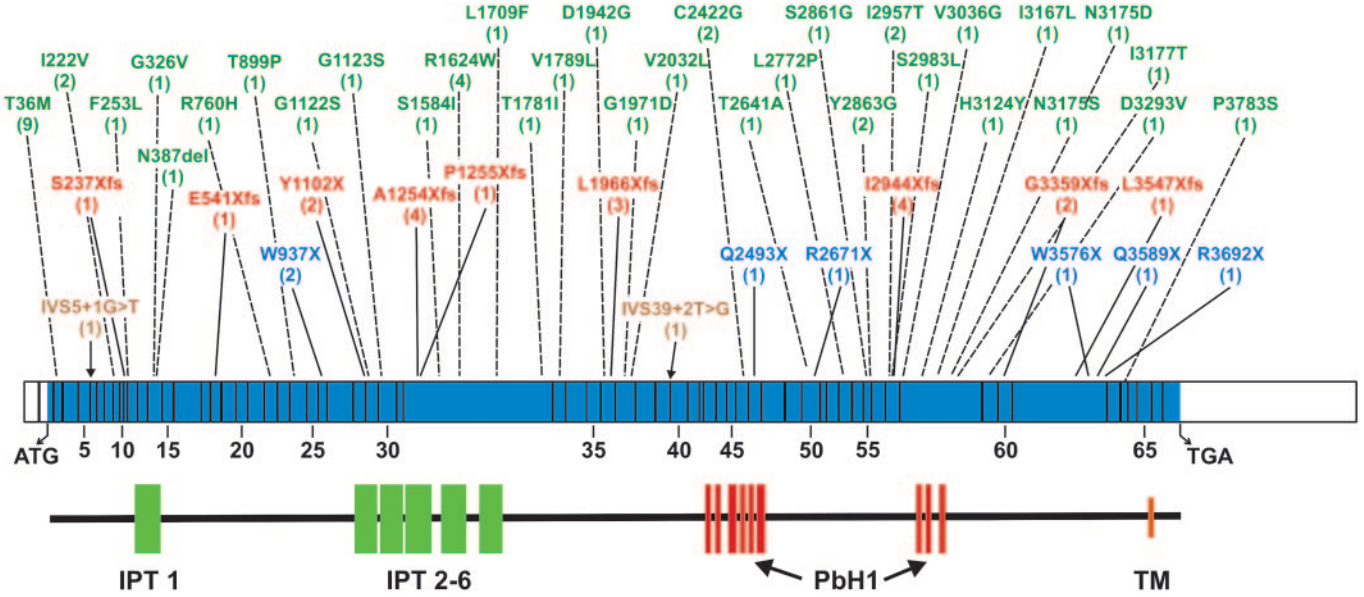

Figure 1. Schematic representation of the location and frequency of mutations in *PKHD1* from our study population. The full-length cDNA composed of the 67 exons of *PKHD1* that encode the longest open reading frame is shown in approximate scale. The locations and nature of the mutations are shown above with the number of occurrences of each mutation in our population shown in parentheses below the mutation name. Amino acid substitutions mutations are shown in green, frame shifting mutations are in red, stop codons are in blue, and splice site mutations are in brown. The approximate domain structure of polyductin/fibrocystin is indicated below.

found (Figure 1), all but three (R760, R1624, and I3177) occurred in residues conserved in the mouse protein (16). The predominance of amino acid substitution mutations is not an artifact of detection because heteroduplex analysis should be most efficient in picking up insertion/deletion mutations resulting in frame shifts and should be equally effective in detecting single nucleotide substitutions regardless of whether they result in amino acid substitutions, nonsense codons, or splice site variants. Because we found mutations distributed roughly evenly along the gene, it is less likely that regional differences in our mutation detection sensitivity lead to underdetection of chain-terminating variants. Furthermore, two other published reports (14,15) found a total of 74 (57%) amino acid substitutions and 55 (43%) chain-terminating mutations on 129 disease chromosomes. Combining all studies, mutations on 204 disease chromosomes were composed of 120 (59%) amino acid substitutions and 84 (41%) putative chain terminations. When the recurrent and overrepresented T36M mutation (38 of 204 disease chromosomes) and the Finnish founder mutation R496X (20 of 204 disease chromosomes (15)) are removed from the calculations, the reports still describe 82 (56%) disease chromosomes with missense changes and 64 (44%) with chain terminations (13–15).

Polymorphism in PKHD1

As part of mutation detection, we defined a spectrum of polymorphisms in *PKHD1* (Table 3). Because the majority of mutations in *PKHD1* are private amino acid substitutions, it is important to develop a catalog of polymorphisms to assist in evaluating gene-based mutation results. We found 14 amino acid substitution polymorphisms that were present in the normal control population. As expected, in contrast to the missense mutations that overwhelmingly occurred in conserved residues (29 of 32; 91% in our study), 7 of 14 (50%) missense polymorphisms occurred in amino acids not conserved in the mouse protein (Table 3). An amino acid substitution previously reported as a mutation (D3139Y) was found to be a polymorphism when an expanded number of normal chromosomes was screened (Table 3) (13,14). The additional intragenic polymorphisms that we found can be used in future haplotype analyses. In the current studies, we did not have access to family members' DNA for the majority of patients and therefore were not able to determine haplotypes for mutant chromosomes.

Genotype-Phenotype Correlations

We sought to determine whether the nature of the underlying mutations played a role in disease severity in ARPKD. We stratified families by clinical presentation into two groups: the severe perinatal group, in which at least one affected child presented with perinatal disease and neonatal demise, and the less severe, nonperinatal group, in which none of the affected children died in the neonatal period. We dichotomized mutations into amino acid substitution and chain-terminating groups—the latter included nonsense, insertion/deletion with frame shifting, and putative splice site variants—on the premise that chain-terminating mutations were likely to be complete loss of function variants, whereas amino acid substitutions may be an admixture of hypomorphic alleles of varying degrees of functional loss (Figure 1). Among families in which at least one mutation was found, 27 families had the severe perinatal presentation and 22 families had nonperinatal presentation. In four families (AL24, AL25, and AL34 in Table 1 and

Table 3. Polymorphisms in *PKHD1*

Nucleotide Change	ORF Change	Exon	Frequency in Controls
Amino acid substitution polymorphisms[a]			
1371G→ C	E457D	16	12/120
2194G→ T	V732F	22	1/120
2278C→ T	R760C	22	7/320
3448G→ C	A1150P	30	1/120
3848C→ T	S1283L	32	2/120
5608T→ G	L1870V	35	1/120
7844C→ G	A2615G	49 (50)	22/320
8606C→ A	T2869K	55 (56)	4/120
9415G→ T	D3139Y	58 (59)	8/320
9428G→ T	R3143I	58 (59)	1/320
10319T→ A	V3440D	61 (65)	3/120
10651G→ A	E3551K	61 (65)	4/120
11696A→ G	Q3899R	66 (70)	18/320
11878G→ A	V3960I	67 (71)	3/320
Silent changes in exons			
Detected in patients, not in controls			
888A→ T		12	0/120
1314G→ A		16	0/120
3393G→ A		30	0/120
5199C→ A		32	0/120
6975C→ T		43 (44)	0/120
7533A→ G		48 (49)	0/120
10299G→ A		61 (65)	0/320
10968A→ G		61 (65)	0 /320
Detected in patients and controls			
214C→ T		4	4/120
237C→ T		4	1/120
276G→ A		4	1/120
1587T→ C		17	1/120
2046A→ C		21	7/120
2214T→ G		22	1/120
2278C→ T		22	3/320
4920A→ C		32	1/320
5896C→ T		36	3/120
7587A→ G		48 (49)	1/120
7764A→ G		49 (50)	5/120
9237G→ A		58 (49)	3/320
9492C→ T		58 (49)	1/320
10521C→ T		61 (65)	2/320
11340T→ C		63 (67)	5/320

Nucleotide Change	Frequency in Controls
Intronic polymorphisms	
IVS3 −36C→ G	1/120
IVS7 +19T→ C	5/120
IVS7 +51G→ T	1/320
IVS8 +67G→ A	1/120
IVS11 −8delT	4/120
IVS15 −10T→ A	1/120
IVS16 −5C→ T	1/120
IVS18 −33C→ G	1/120
IVS32 +14A→ G	5/120
IVS32 −32G→ A	3/320
IVS35 −42T→ G	5/320
IVS38 (38) −7insC	7/320
IVS47 (48) −4T→ C	4/320
IVS51 (52) +12T→ A	21/320
IVS52 (53) −32G→ C	2/120
IVS53 (54) −54G→ T	4/320
IVS54 (55) −73C→ T	9/320
IVS54 (55) −24A→ G	4/320
IVS55 (56) −32insC	7/320
IVS55 (56) −19A→ C	1 /60
IVS55 (56) +44A→ C	5/160
IVS56 (57) +62G→ A	11/160
IVS56 (57) +44A→ C	2 /60
IVS58 (59) +10T→ G	6/160
IVS61 (64) +48insG	5/160

[a] Naming conventions and exon numbering as in Table 1.

306/1272 in Onuchic *et al.* (13)), we did not have enough clinical information, usually because of prenatal diagnosis and termination, for inclusion in the analysis. Among the 75 disease chromosomes from the 53 families for which we identified the mutations, 37 were from families with the severe perinatal form, 32 were from families with the nonperinatal disease, and 6 were from families with unknown disease severity (Table 4).

We found that amino acid substitution mutations were more commonly associated with the nonperinatal phenotypes, whereas chain-terminating mutations were more commonly associated with the perinatal severe phenotype ($\chi^2 = 7.89$, $P = 0.005$; Table 4). When we combine our study with published studies by Bergmann *et al.* (15) and Ward *et al.* (14), the association of amino acid substitutions with nonperinatal presentation and of chain-termination mutations with perinatal presentations in 197 disease chromosomes remains highly significant ($\chi^2 = 19.65$, $P = 9.4 \times 10^{-6}$; Table 4). Excluding both T36M and R496X to eliminate potential bias introduced by overrepresented mutations does not alter this significant association among the remaining 140 disease chromosomes ($\chi^2 = 11.24$, $P = 8.1 \times 10^{-4}$). We further examined these data by evaluating genotype and phenotype in 20 patients in whom we found the mutations on both disease chromosomes. We found a trend toward association of amino acid substitutions with milder presentations and chain terminations with more severe presentation ($\chi^2 = 3.96$, $P = 0.047$; Fisher exact $P = 0.091$). When we combined these data with the studies of Bergmann *et al.* (15) and Ward *et al.* (14), the association of mutations with severity of phenotype in individuals in whom both mutations were known was confirmed ($n = 68$ patients; 136 chromosomes; $\chi^2 = 9.08$, $P = 0.0026$; Table 4).

Finally, we examined the correlations of phenotype in patients with two amino acid substitution mutations, two chain-terminating mutations, or one of each (Table 4). When we combined our 20 patients with both mutations known with those of Bergmann *et al.* (15) and Ward *et al.* (14) ($n = 68$), we found that all patients who survived the perinatal period ($n = 24$) had at least one amino acid substitution. Conversely, all patients with two predicted chain-termination mutations ($n = 16$) had perinatal presentation with neonatal death. Thus, the presence of a single amino acid substitution is required for presentation of a nonperinatal phenotype ($\chi^2 = 11.54$, $P = 0.003$). Having either one or two amino acid substitution mutations did not preclude a severe perinatal phenotype, although the trend was toward a milder presentation. The data are consistent with the notion that the nature of the germline mutations in ARPKD play a role in determining the phenotype and that amino acid substitution mutations can be less severe than mutations that lead to chain termination.

Table 4. Mutation type and phenotype in *PKHD1*

Phenotype	Mutation Type[a]			Mutation Type[b]		
	Amino Acid Substitution	Chain-Terminating	Total	Amino Acid Substitution	Chain-Terminating	Total
Nonperinatal						
Actual	26	6	32	65	19	84
Expected	20.8	11.2	32	49.9	34.3	84
Perinatal						
Actual	18	19	37	52	61	113
Expected	23.6	13.4	37	67.1	45.9	113
Total	44	25	69	117	80	197

Phenotype	Mutation Type[c]			Combination of Mutation Type[d]			
	Amino Acid Substitution	Chain-Terminating	Total	Missense/Missense	Missense/Chain-Termination	Chain-Terminating/Chain-Terminating	Total
Nonperinatal							
Actual	36	14	50	10	14	0	24
Expected	27.6	22.4	50	8.1	10.2	5.7	24
Perinatal							
Actual	39	47	86	13	15	16	44
Expected	47.4	38.6	86	14.9	18.8	10.3	44
Total	75	61	136	23	29	16	68

[a] All mutations in the current study and (13) where the clinical phenotype was known; $\chi^2 = 7.89$, $P = 0.005$.
[b] All mutations in the current study and (13–15) where the clinical phenotype was known; $\chi^2 = 19.65$, $P = 9 \times 10^{-6}$.
[c] All mutations in the current study and (13–15) where both mutations in an individual as well as the clinical phenotype was known; $\chi^2 = 9.08$, $P = 0.0026$.
[d] All mutations in the current study and (13–15) where both mutations in an individual as well as the clinical phenotype was known; $\chi^2 = 11.54$, $P = 0.003$.

Discussion

An expected benefit from human disease gene identification is the ability to understand the molecular basis of clinical variation in disease presentation and outcome. Such correlation of the underlying molecular defect with the clinical course can improve both the care provided to patients and the understanding of the molecular pathogenesis of the disease. Historically, ARPKD has been subdivided clinically into four distinct phenotypes on the basis of the age at presentation and the proportion of dilated collecting ducts (6). The degree of hepatic fibrosis has been noted to be inversely correlated to both the age at presentation and the degree of cystic involvement of the kidneys, but significant variation in this relationship has been described (7,8). There is relative concordance in the clinical presentation of ARPKD among siblings in most cases (17,18). There are, however, numerous instances of siblings that are discordant across the spectrum of renal and biliary tract involvement (17,19). Nonetheless, it is likely that allelic variation plays a significant role in the interfamilial variation that is well described for the disease presentation (8,7,20). Such allelism is likely further modified by either additional genetic or environmental factors that account for intrafamilial variation and contribute to the interfamilial variation in the setting of comparable genotypes.

We examined the contribution of germline mutations to the ARPKD phenotype. We used a binary assessment of clinical presentation based on whether the affected children survived the immediate perinatal period. This end point was chosen because it was reliably available for almost all of our patient sample and could be gleaned from the recently published literature reporting additional mutations in ARPKD. In addition, numerous studies have suggested that the stratified classification of Blyth and Ockenden (6) has limited clinical utility and is not genetically relevant (reviewed in (2)). The hypothesis that we sought to test was whether at least part of the difference in the severity of clinical presentation between families results from the nature of the germline mutations. We found that in our patient sample and in those published with known mutations, survival past the immediate perinatal period required the presence of at least one amino acid substitution mutation; the presence of two chain-terminating mutations invariably resulted in perinatal lethality. However, the converse did not apply: the presence of an apparent missense variant did not guarantee escape from neonatal lethality. One interpretation of the finding of presence of missense mutations is required for but not necessarily sufficient for escape from neonatal lethality is that some amino acid substitutions act as hypomorphic alleles with reduced function, whereas others result in complete loss of function similar to chain-terminating mutations. Phenotypic variation in the effects of the hypomorphic amino acid substitutions are further determined by genetic or environmental modifiers; such modification of complete loss of function alleles is less likely.

Because most mutations recur infrequently, it so far has not been possible to get sufficient numbers of specific variants in unrelated patients to draw significant conclusions regarding direct correlation of individual mutations with severity. Inspection of the majority of recurrent missense variants did not allow conclusion that particular amino acid substitutions were exclusively associated with perinatal survival. Some recurrent mutations (*e.g.*, I222V, I2331K, R1624W, Y2863G) were more often found in milder presentations, but the number of chromosomes for each was small. The T36M variant is unique in this regard because it is recurrent and seems to have arisen *de novo* several times. Among 68 patients with two known mutations, T36M was found on 12 chromosomes in 11 patients. Among these 11 patients, 6 had T36M in conjunction with another amino acid substitution and 1 was homozygous for T36M (15). Only 2 of 11 patients, with T36M in association with I222V and V3471G, had the less severe presentation. The remainder had the severe perinatal presentation, including all four patients with T36M in conjunction with chain-terminating mutations and the one patient homozygous for T36M. It is interesting that the T36M amino acid substitution represents a potential alternative initiation codon that is actually predicted to be stronger than the native start codon when analyzed by the NetStart program (www.cbs.dtu.dk/services/NetStart). If the alternative start site were used exclusively *in vivo*, then the translation product would lack the leader sequence required for proper folding and T36M may indeed behave like a complete loss of function allele akin to the chain-terminating mutations. The clinical findings reported to date are consistent with this hypothesis, but additional studies, including functional assessment of T36M, are required for verification.

It is possible that the location of amino acid substitutions along the polyductin/fibrocystin protein may influence the severity of the resultant phenotype. Our detection of nucleotide substitutions was fairly homogeneous across the gene (Figure 1), and the analysis of Bergmann *et al.* (15) also seemed well distributed across the gene. Among the 68 patients with two known mutations, there were 39 different amino acid substitution mutations accounting for 75 of 136 disease chromosomes. Of the 39 different missense changes, 17 occurred exclusively on chromosomes in nonneonatal lethal cases, whereas 22 occurred on chromosomes associated with severe perinatal cases; some of the latter were also found in nonneonatal lethal presentations. The distribution of amino acid substitutions found exclusively in less severely affected individuals was 3, 9, 3, and 2 in successive 1000 amino acids quartiles of the 4074 amino acid protein. The distribution of missense variants associated with neonatal lethality was 8, 5, 5, and 4 in the respective quartiles. More than 50% (9 of 17) of amino acid substitutions in milder presentations occurred between residues 1000 and 2000; of these, 6 (35%) of 17 occurred in 10% of the sequence between amino acids 1600 and 2000. Thus, one third of missense variants exclusively associated with milder presentation occurred in a region of the protein between the clusters of IPT domains and PbH1 repeats that is devoid of identifiable structural motifs (Figure 1). Further studies with larger numbers of mutations defined in clinically well-characterized patient populations will better define the significance, if any, of the location or the type of amino acid substitution mutations with respect to disease prognosis.

Using the disease severity stratification criteria that we chose, we were unable to see a trend toward better outcomes with two, as opposed to one, amino acid substitution mutations. There are several potential explanations for this. The first is that we used a binary outcome measure (neonatal survival *versus* no survival), and it is possible that to see the mutation "dosage" effect of having one or two missense variants, a more continuous outcome measure (*e.g.*, length of survival to kidney failure or liver transplantation) may be required. Another possibility is that a significant proportion (*e.g.*, >50%) of amino acid substitution mutations are pathophysiologically equivalent to chain-termination mutations. Under these circumstances, a much larger group of patients with two known mutations may be needed to see the dosage effect. The third consideration is the effect of modifying factors. In light of the several reports of marked intrafamilial variation in ARPKD (17,19) and our observations that certain mutations can be associated with severe as well as less severe outcomes, it is likely that gene-by-gene and gene-by-environment interactions play an additive role in outcomes of patients. The uniformly severe outcome when patients have two chain-terminating mutations support the hypothesis that modifiers will likely have their greatest effect in the setting of one or two amino acid substitution mutations that have some degree of residual function that allow for a less severe presentation.

Finally, it has been noted that the primary cause of neonatal lethality in ARPKD is pulmonary hypoplasia in the setting of massive enlargement of the kidneys and intrauterine renal failure with consequent oligohydramnios (reviewed in (2,21)). One way to reconcile these clinical observations with the finding that apparent complete loss of function mutations are associated with the severe presentation is to hypothesize that the kidney phenotype varies with the "molecular severity" of the mutation and in turn brings about early death through pulmonary failure. If the molecular lesion is milder, then both kidney and liver diseases run their relative courses, leading to the variable prognosis of children's surviving the neonatal period. There is precedent for the kidney tubules being developmentally more sensitive to polycystic disease mutations than the bile duct. Homozygous inactivation of either *Pkd1* or *Pkd2*, the dominant polycystic kidney disease orthologs, results in severe intrauterine renal cystic disease with normal-appearing livers (22,23). However, when embryonic lethality is circumvented, surviving mice develop bile duct–derived cysts in addition to massive kidney cysts in postnatal life (S. Somlo, unpublished observations). Thus, it may be that early in development, the kidney is most sensitive to loss of *PKHD1*. Manifestation of the severe clinical liver lesion requires survival to a later stage. An extreme example of this is the two patients in our study (AL31 and AL33) who had portal hypertension but no discernible kidney disease from either clinical or radiologic perspectives. The known mutation in each of these patients was a unique amino acid substitution (Table 1). Understanding the genetic factors underlying the clinical course of ARPKD will likely be the first step in tailoring clinical management of ARPKD to individual families' needs.

Acknowledgments

This work was supported by National Institutes of Health Grant R01 DK51259.

The authors thank the many families and their health care providers for their cooperation with these studies, Richard Lifton for helpful discussions, and Emily Hilton and the HHMI Biopolymer–Keck Foundation Biotechnology Resource Laboratory for technical assistance with sequencing.

References

1. Zerres K, Mucher G, Becker J, Steinkamm C, Rudnik-Schoneborn S, Heikkila P, Rapola J, Salonen R, Germino GG, Onuchic L, Somlo S, Avner ED, Harman LA, Stockwin JM, Guay-Woodford LM: Prenatal diagnosis of autosomal recessive polycystic kidney disease (ARPKD): Molecular genetics, clinical experience, and fetal morphology. *Am J Med Genet* 76: 137–144, 1998
2. Guay-Woodford LM: Autosomal recessive polycystic kidney disease: Clinical and genetic profiles. In: *Polycystic Kidney Disease*, edited by Watson ML, Torres VE, Oxford, Oxford University Press, 1996, pp 237–267
3. Fonck C, Chauveau D, Gagnadoux MF, Pirson Y, Grunfeld JP: Autosomal recessive polycystic kidney disease in adulthood. *Nephrol Dial Transplant* 16: 1648–1652, 2001
4. Zerres K, Mucher G, Bachner L, Deschennes G, Eggermann T, Kaariainen M, Knapp T, Lennert T, Misselwitz J, von Muhlendahl KE, Neumann HPH, Pirson Y, Rudnick-Schoneborn S, Steinbicker V, Wirth B, Scharer K: Mapping of the gene for autosomal recessive polycystic kidney disease (ARPKD) to chromosome 6p21-cen. *Nat Genet* 7: 429–432, 1994
5. Guay-Woodford LM, Muecher G, Hopkins SD, Avner ED, Germino GG, Guillot AP, Herrin J, Holleman R, Irons DA, Primack W: The severe perinatal form of autosomal recessive polycystic kidney disease maps to chromosome 6p21.1-p12: Implications for genetic counseling. *Am J Hum Genet* 56: 1101–1107, 1995
6. Blyth H, Ockenden BG: Polycystic disease of kidney and liver presenting in childhood. *J Med Genet* 8: 257–284, 1971
7. 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. *Nephrol Acta Paediatr* 85: 437–445, 1996
8. Kaplan BS, Fay J, Shah V, Dillon MJ, Barratt TM: Autosomal recessive polycystic kidney disease. *Pediatr Nephrol* 3: 43–49, 1989
9. Gang DL, Herrin JT: Infantile polycystic disease of the liver and kidneys. *Clin Nephrol* 25: 28–36, 1986
10. Holthofer H, Kumpulainen T, Rapola J: Polycystic disease of the kidney. Evaluation and classification based on nephron segment and cell-type specific markers. *Lab Invest* 62: 363–369, 1990
11. Roy S, Dillon MJ, Trompeter RS, Barratt TM: Autosomal recessive polycystic kidney disease: Long-term outcome of neonatal survivors. *Pediatr Nephrol* 11: 302–306, 1997
12. Desmet VJ: Congenital diseases of intrahepatic bile ducts: Variations on the theme "ductal plate malformation." *Hepatology* 16: 1069–1083, 1992
13. 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* 70: 1305–1317, 2002
14. 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. *Nat Genet* 30: 259–269, 2002
15. Bergmann C, Senderek J, Sedlacek B, Pegiazoglou I, Puglia P, Eggermann T, Rudnik-Schneborn 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* 14: 76–89, 2003
16. Nagasawa Y, Matthiesen S, Onuchic LF, Hou X, Bergmann C, Esquivel E, Senderek J, Ren Z, Zeltner R, Furu L, Avener ED, Moser M, Somlo S, Guay-Woodford LM, Buttner R, Zerres K, Germino GG: Identification and characterization of *Pkhd1*, the mouse orthologue of the human ARPKD gene. *J Am Soc Nephrol* 13: 2246–2258, 2002
17. Deget F, Rudnik-Schoneborn S, Zerres K: Course of autosomal recessive polycystic kidney disease (ARPKD) in siblings: A clinical comparison of 20 sibships. *Clin Genet* 47: 248–253, 1995
18. Zerres K, Rudnik-Schoneborn S, Deget F: Childhood onset autosomal dominant polycystic kidney disease in sibs: Clinical picture and recurrence risk. German Working Group on Paediatric Nephrology (Arbeitsgemeinschaft fur Padiatrische Nephrologie). *J Med Genet* 30: 583–588, 1993
19. Kaplan BS, Kaplan P, de Chadarevian JP, Jequier S, O'Regan S, Russo P: Variable expression of autosomal recessive polycystic kidney disease and congenital hepatic fibrosis within a family. *Am J Med Genet* 29: 639–647, 1988
20. Herrin JT: Phenotypic correlates of autosomal recessive (infantile) polycystic disease of kidney and liver: Criteria for classification and genetic counseling. *Prog Clin Biol Res* 305: 45–54, 1989
21. Capisonda R, Phan V, Traubuci J, Daneman A, Balfe JW, Guay-Woodford LM: Autosomal recessive polycystic kidney disease: Outcomes from a single-center experience. *Pediatr Nephrol* 18: 119–126, 2003
22. Lu W, Peissel B, Babakhanlou H, Pavlova A, Geng L, Fan X, Larson C, Brent G, Zhou J: Perinatal lethality with kidney and pancreas defects in mice with a targetted Pkd1 mutation. *Nat Genet* 17: 179–181, 1997
23. Wu G, Markowitz GS, Li L, D'Agati VD, Factor SM, Geng L, Tibara S, Tuchman J, Cai Y, Park JH, van Adelsberg J, Hou H Jr, Kucherlapati R, Edelmann W, Somlo S: Cardiac defects and renal failure in mice with targeted mutations in Pkd2. *Nat Genet* 24: 75–78, 2000