Journal of Human Genetics (2016) 61, 811–821

www.nature.com/jhg

ORIGINAL ARTICLE

Expanding the mutation spectrum in 130 probands with ARPKD: identification of 62 novel *PKHD1* mutations by sanger sequencing and MLPA analysis

Salvatore Melchionda[1,5], Teresa Palladino[1,5], Stefano Castellana[2], Mario Giordano[3], Elisa Benetti[4], Patrizia De Bonis[1], Leopoldo Zelante[1] and Luigi Bisceglia[1]

Autosomal recessive polycystic kidney disease (ARPKD) is a rare severe genetic disorder arising in the perinatal period, although a late-onset presentation of the disease has been described. Pulmonary hypoplasia is the major cause of morbidity and mortality in the newborn period. ARPKD is caused by mutations in the *PKHD1* (*polycystic kidney and hepatic disease 1*) gene that is among the largest human genes. To achieve a molecular diagnosis of the disease, a large series of Italian affected subjects were recruited. Exhaustive mutation analysis of *PKHD1* gene was carried out by Sanger sequencing and multiple ligation probe amplification (MLPA) technique in 110 individuals. A total of 173 mutations resulting in a detection rate of 78.6% were identified. Additional 20 unrelated patients, in whom it was not possible to analyze the whole coding sequence, have been included in this study. Taking into account the total number ($n = 130$) of this cohort of patients, 107 different types of mutations have been detected in 193 mutated alleles. Out of 107 mutations, 62 were novel: 11 nonsense, 6 frameshift, 7 splice site mutations, 2 in-frame deletions and 2 multiexon deletion detected by MLPA. Thirty-four were missense variants. In conclusion, our report expands the spectrum of *PKHD1* mutations and confirms the heterogeneity of this disorder. The population under study represents the largest Italian ARPKD cohort reported to date. The estimated costs and the time invested for molecular screening of genes with large size and allelic heterogeneity such as *PKHD1* demand the use of next-generation sequencing (NGS) technologies for a faster and cheaper screening of the affected subjects.
Journal of Human Genetics (2016) **61,** 811–821; doi:10.1038/jhg.2016.58; published online 26 May 2016

INTRODUCTION

Autosomal recessive polycystic kidney disease (ARPKD; Online Mendelian Inheritance in Man (OMIM) 263200) is a rare severe genetic disorder that occurs in 1:20 000 live births.[1] Typically, the disease arises in the perinatal period and patients present with enlarged kidneys and liver, respiratory failure, hypertension and urinary tract infection. Moreover, enlarged and echogenic kidneys, leading to oligohydramnios and possibly to pulmonary hypoplasia, can be observed *in utero* by fetal ultrasound. Pulmonary hypoplasia is the major cause of morbidity and mortality in the newborn period. In the survivors, hypertension and renal insufficiency, including end-stage renal disease (up to one-third of children require renal replacement therapy), are the major signs of renal disease.[2] A further pathognomonic sign of disease is the biliary dysgenesis resulting in congenital hepatic fibrosis, plus intrahepatic bile duct dilatation (Caroli disease). Late-onset presentation of the disease, in childhood or adulthood, has been reported, and these patients predominantly have liver disease and milder kidney involvement.[3,4] ARPKD is caused by mutations of the *PKHD1* (*polycystic kidney and hepatic disease 1*) gene mapped on the short arm of chromosome 6 (6p12.2).[5,6] *PKHD1* is one of the largest human genes, extending over 469 kb encompassing 86 exons, 67 of which lead to the longest transcript encoding a 4074-amino-acid protein, polyductin/fibrocystin. The protein has a predicted molecular weight of 447 kDa and it is expressed in the basal body and primary cilia of renal and bile duct epithelial cells.[6,7] Mutations are spread along the entire gene, but they are not equally scattered. Except for a few population-specific founder alleles (that is, p.Arg496*,[8] p.Met627Lys[9] and the common p.Thr36Met amino acid change), *PKHD1* is characterized by significant allelic heterogeneity and the great majority of the patients are compound heterozygous. The screenings of the gene have been performed on several cohorts of different sample size as well as on isolated familial cases. To date, >800 variants have been identified in the *PKHD1* gene, 748 of which have been recorded in the public ARPKD/*PKHD1* database

[1]Medical Genetics Unit, IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy; [2]Bioinformatics Unit, IRCCS Casa Sollievo della Sofferenza-Mendel, Rome, Italy; [3]Pediatric Nephrology and Dialysis Unit, Pediatric Hospital Giovanni XXIII, Bari, Italy and [4]Pediatric Nephrology, Dialysis and Transplant Unit, Women's and Children's Health Department, University of Padua, Padua, Italy
[5]These two authors contributed equally to this work.
Correspondence: Dr L Bisceglia, Medical Genetics Unit, IRCCS Casa Sollievo della Sofferenza Hospital, Poliambulatorio Giovanni Paolo II, San Giovanni Rotondo, FG I-71013, Italy.
E-mail: l.bisceglia@operapadrepio.it
Received 21 December 2015; revised 15 April 2016; accepted 28 April 2016; published online 26 May 2016

(http://www.humgen.rwth-aachen.de; last update: August 2013).[8–32] Detection rates ranging between 42 and 87% depend on the technique used and the heterogeneity of the population studied. The mutational spectrum consists of truncating mutations (frameshift, nonsense and splice site) and a large number of missense mutations spread over the entire gene. Single-exon–multiexon deletions/duplications of the *PKHD1* gene are very rare, with these copy number variations being reported only by three authors.[16,22,32] In individuals who are *PKHD1* negative, mutations in other PKD genes, such as *HNF1B*, *PKD1* and *PKD2*, have been detected.[33–35] In 2005, based on the analysis of a large number of *PKHD1*-mutated alleles reported in literature, Bergman *et al.*[36] defined a worldwide exon algorithm for *PKHD1* mutation screening, with 50% of the detected mutations affecting 7 exons (3, 58, 32, 36, 57, 61 and 9). Taking into account the algorithm proposed by Bergmann *et al.*,[36] we analyzed by Sanger sequencing 130 unrelated ARPKD patients, 117 of whom were of Italian origin. In addition, we screened a subset of the patients by applying the multiple ligation probe amplification (MLPA) technique in order to search for exon deletions/duplications.

MATERIALS AND METHODS

Patients

Up to December 2013, at Medical Genetics Center of IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo (Italy), we collected blood samples from 116 unrelated ARPKD patients and 28 parents of 14 not available affected subjects (59 females and 71 males). Seventy-six showed a disease onset at <1 year of age (27 were neonatal cases, 40 perinatal and 9 were pregnancy termination cases) and 54 presented a childhood, juvenile or adult disease onset with an average age of 13.3 years (range 1–54). For each individual, a signed informed consent for genetic testing was obtained.

Inclusion criteria for an ARPKD diagnosis were dilated collecting ducts and congenital hepatic fibrosis accordingly with an available renal and liver histology. If the histology was not available, imaging findings should include strong evidence of ARPKD (bilateral nephromegaly, increased renal echogenicity, loss of corticomedullar differentiation and signs of hepatic fibrosis). Moreover, all probands presented with at least one of the following clinical features: oligohydramnios or anhydramnios, pulmonary hypoplasia, portal hypertension, Potter's face or affected siblings.

This study was approved by the Ethics Committee of the IRCCS CSS Hospital and complies with the guidelines of the Declaration of Helsinki.

DNA amplification and sequencing

DNA was extracted from blood using EZ1 DNA Blood Kit (QIAGEN, Hilden, Germany). PCR was performed with 70–80 ng of genomic DNA, 15 pmol of each primer and AmpliTaq Gold DNA polymerase (Applied Biosystems, Austin, TX, USA) in the following conditions: 12 min at 94 °C followed by 35 cycles of steps at 94 °C, 56–58–60–62–64–65 °C and 72 °C. PCR products were purified with ExoSAP-IT (USB, Affymetrix, Cleveland, OH, USA) and sequenced in both directions using BigDye terminator v1.1 chemistry on an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems, Austin, TX, USA).

DNA sequencing was performed on all coding exons (2–67) of the longest open reading frame of *PKHD1* and their intronic flanking sequences (20 to 70 bp for each side of exon). These regions were amplified in 74 amplicons. Exons 32, 58 and 61 were sequenced in overlapping fragments because of their large size (5, 3 and 3 fragments, respectively). Primers were designed using the primer 3 program (http://frodo.wi.mit.edu/) and the amplicons were arranged and analyzed according first to the algorithm proposed by Bergmann *et al.*[36] and second to the annealing temperatures in order to optimize the workflow. DNA alignment with the reference sequence (NM_138694.3) and variant analysis has been carried out using Sequencher (GeneCodes, Ann Arbor, MI, USA).

MLPA analysis

Patients resulting negative or carrying a single *PKHD1* mutation underwent MLPA analysis in order to detect gene deletions or duplications. Similarly, subjects carrying variants with *a priori* doubt pathogenicity (that is, novel missense variants) were also analyzed by MLPA.

P341 and P342 probemix developed by R Vijzelaar at MRC-Holland (Amsterdam, The Netherlands) containing 71 probes, interrogating all exons of the longest open reading frame (NM_138694.3 sequence), except for exon 17, were used.

MLPA was performed according to the manufacturer's protocol. Briefly, 100 ng of genomic DNA was used as starting material; after hybridization, ligation and amplification, the PCR products were size-separated by capillary electrophoresis on an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Electropherograms were visualized by GeneMapper v.4.0 (Applied Biosystems, Foster City, CA, USA) and data analysis was carried out using the Coffalyser.Net, the free software designed specifically by MRC-Holland for the analysis of MLPA data and compatible with data files produced directly by all major capillary electrophoresis systems.

Samples showing copy number variations at MLPA analysis were reanalyzed in a second independent experiment and, when available, the result has been confirmed in the parents and other available relatives.

Bioinformatics analysis

The putative pathogenicity of nonsynonymous *PKHD1* variants was inspected by querying the dbNSFP version 3.0b2a,[37] a database containing multiple precomputed functional impact predictions for every possible amino acid changing mutation within the human genome. In particular, SIFT,[38] PROVEAN 1.1,[39] PolyPhen-2 v2.2.2,[40] LRT,[41] MutationTaster2,[42] MutationAssessor2,[43] FATHMM v2.3,[44] and CADD v1.2[45] predictions were extracted for each single missense variant.

These tools implement different methods (often based on supervised machine learning) in order to assign a pathogenicity score together with a (typically) double categorical predictions ('neutral' or 'pathogenic').

Furthermore, evolutionary conservation on corresponding genomic sites was inspected by retrieving the PhyloP7way[46] and GERP++ indexes[47] (see Table 3 for details).

Mutations were also checked for their presence in public variant databases, such as dbSNPv144[48] and ExAC v0.3 (Exome Aggregation Consortium; http://exac.broadinstitute.org/, accessed 20 October 2015).

Regarding *PKHD1* variants located in the splicing regions, the possible outcome was predicted by using the Human Splice Finder algorithm on the HSF3.0 webserver,[49] a position weight matrix-based method that identifies splicing signals and determines a similarity score variation between the wild-type and the mutant one with respect to a signal-specific position weight matrix. Briefly, splice site signal is identified along a DNA sequence for its similarity with a matrix (that is, it reaches a determined score threshold); if a mutation alters such sequence, the splice site is tagged as 'broken' when its score decrease for at least 10% of the wild-type value. Conversely, when similarity score increases for 10%, then a new putative splice site has been created by the considered mutation.

Furthermore, information on protein structure and domains was retrieved by querying the Uniprot (http://www.uniprot.org/) and the PFam web resources (http://pfam.xfam.org/).

Table 1 Number of patients characterized for *PKHD1* mutations by exhaustive sequencing and MLPA analysis

Patients	*Italian (n = 97)*	*Non-Italian (n = 13)*	*Total (%) (n = 110)*
Homozygous	5	5	75 (68.2%)
Compound heterozygous	62	3	
Heterozygous	20	3	23 (20.9%)
Negatives	10	2	12 (10.9%)

Abbreviations: MLPA, multiple ligation probe amplification; *PKHD1*, polycystic kidney and hepatic disease 1.

Table 2 Genotype description in a cohort of 130 ARPKD families

Patient	*Ethnic origin*	*Age at enrollment/death*	*PKHD1 mutation*	*Type of mutation*	*Exon*	*Parent analysis*
pkd1	I	3 y	p.Ala1758Gly (c.5273C>T)	Missense [n]	33	NA
			c.5277delA (p.Ser1761Leufs*42)	Frameshift [n]	33	
pkd2	I	22 y	p.Arg2671* (c.8011C>T)	Stop codon	50	NA
			p.Gly466Glu (c.1397G>A)	Missense	16	
pkd12	I	Pregnancy termination	p.Ser1929Pro (c.5785T>C)	Missense [n]	36	Y2
			c.5895dupA (p.Leu1966Thrfs*4)	Frameshift	36	
pkd15	I	4 y	p.Thr36Met (c.107C>T)	Missense	3	Y2
			p.Asn901Thr (c.2702A>C)	Missense	25	
pkd16	I	4 y	p.Thr36Met (c.107C>T)	Missense	4	Y2
			c.7912-2A>G	Splicing [n]	IVS49	
pkd24	I	16 y	p.Ile2957Thr (c.8870T>C)	Missense	57	NA
			p.Arg1624Trp (c.4870C>T)	Missense	32	
pkd25	I	Neonatal demise	p.Gln795* (c.2383C>T)	Stop codon [n]	23	PP
			p.Pro1628Ala (c.4882C>G)	Missense	32	
pkd28	I	10 y	p.Arg496* (c.1486C>T)	Stop codon	16	NA
			p.Ser3524Phe (c.10571C>T)	Missense [n]	61	
pkd30	I	24 y	p.Asp2528Asn (c.7582G>A)	Missense [n]	48	NA
			p.Ala3224Pro (c.9670G>C) (hom)	Missense [n]	58	
pkd31	I	23 y	p.Ile2957Thr (c.8870T>C)	Missense	57	NA
			p.Ala3224Pro (c.9670G>C)	Missense [n]	58	
pkd45	I	Neonatal demise	p.Arg375Trp (c.1123C>T)	Missense	15	Y2
			p.Trp848* (c.2543G>A)	Stop codon [n]	24	
pkd51	C	2 m	p.Arg3772* (c.11314C>T)	Stop codon	63	Y2
			p.Arg3772* (c.11314C>T)	Stop codon	63	
pkd52	I	Perinatal (alive)	p.Ser1929Pro (c.5785T>C)	Missense [n]	36	Y2
			p.Met1? (c.1A>G)	Missense [n]	2	
pkd61	I	Perinatal (alive)	p.Thr36Met (c.107C>T)	Missense	8	Y2
			c.5751+3G>A	Splicing	IVS35	
pkd62	I	Neonatal demise	p.Arg723His (c.2168G>A)	missense [n]	22	PP
			p.Ser2639* (c.7916C>A)	Stop codon	50	
pkd67	I	6 m	p.Gly466Glu (c.1397G>A)	Missense	16	Y2
			p.Cys3394Arg (c.10180T>C)	Missense [n]	61	
pkd72	A	15 y	p.Glu1448Gly (c.4343A>G)	Missense #	32	NA
			p.Glu1448Gly (c.4343A>G)	Missense #	32	
pkd78	I	2 m	p.Gln2304* (c.6910C>T)	Stop codon [n]	43	Y2
			p.Arg2573Cys (c.7717C>T)	Missense	48	
pkd82	I	Pregnancy termination	c.5895dupA (p.Leu1966Thrfs*4)	Frameshift	36	Y2
			c.3365-3C>T	Splicing [n]	IVS29	
pkd85	I	Perinatal (alive)	p.Asn1779Ser (c.5336A>G)	Missense [n]	33	Y2
			p.Val3263Ala (c.9788T>C)	Missense	58	
pkd90	I	4 y	p.Glu1202Lys (c.3604G>A)	Missense [n]	31	Y2
			p.Ser2861Gly (c.8581A>G)	Missense	55	
pkd92	I	Neonatal demise	p.Ser3570* (c.10709C>G)	Stop codon	61	PP
			c.11151_11153delGTC (p.Ser3718del)	In frame [n]	61	
pkd93	I	Neonatal demise	p.Arg1775* (c.5323C>T)	Stop codon	33	Y2
			p.Arg1775* (c.5323C>T)	Stop codon	33	
pkd95	I	Perinatal (alive)	p.Pro1628Ala (c.4882C>G)	Missense	32	Y2
			c.9523_9524delinsG (p.Asn3175Valfs*8)	Frameshift [n]	58	
pkd97	I	Neonatal demise	p.Arg760His (c.2279G>A)	Missense	22	PP
			p.Arg760His (c.2279G>A)	Missense	22	
pkd99	I	4 y	p.Arg1624Trp (c.4870C>T)	Missense	32	Y2
			p.Arg3482Cys (c.10444C>T)	Missense	61	
pkd101	I	7 m	p.Ile222Val (c.664A>G)	Missense	9	Y2
			p.Val2921Gly (c.8762T>G)	Missense [n]	56	
pkd102	I	Pregnancy termination	p.Val2921Gly (c.8762T>G)	Missense [n]	56	Y2
			c.2140+1G>A	Splicing [n]	IVS21	
pkd107	I	Neonatal demise	p.Gln2824* (c.8470C>T)	Stop codon [n]	54	Y2
			p.Asn1779Ser (c.5336A>G)	Missense [n]	33	
pkd108	I	Neonatal demise	c.1854delA (p.Gly619Alafs*3)	Frameshift	20	Y2
			c.1854delA (p.Gly619Alafs*3)	Frameshift	20	

Table 2 (Continued)

Patient	*Ethnic origin*	*Age at enrollment/death*	*PKHD1 mutation*	*Type of mutation*	*Exon*	*Parent analysis*
pkd111	I	4 y	p.Ser1156Leu (c.3467C>T)	Missense	30	NA
			c.6122-?_6490+?del (p.Ser2042_Gly2164del)	Large deletion [n]	38–39	
pkd116	I	12 y	p.Asp1944Asn (c.5981G>C)	Missense [n]	36	Y2
			p.Gln2708* (c.8122C>T)	Stop codon [n]	51	
pkd117	I	8 y	p.Met1? (c.1A>G)	Missense [n]	2	Y2
			p.Ile222Val (c.664A>G)	Missense	9	
pkd118	I	Neonatal demise	p.Gln1473* (c.4417C>T)	Stop codon [n]	32	Y2
			c.5895dupA (p.Leu1966Thrfs*4)	Frameshift	36	
pkd119	I	Neonatal demise	p.Asn2300Lys (c.6900C>G)	Missense [n]	43	PP
			p.Asn2300Lys (c.6900C>G)	Missense [n]	43	
pkd120	I	3 m	c.5895dupA (p.Leu1966Thrfs*4)	Frameshift	36	NA
			p.Gln2708* (c.8122C>T)	Stop codon [n]	51	
pkd125	I	Neonatal demise	p.Thr36Met (c.107C>T)	Missense	3	Y2
			p.Ser2861Gly (c.8581A>G)	Missense	55	
pkd126	I	11 m	p.Ile2331Lys (c.6992T>A)	Missense	43	Y2
			p.Gln2708* (c.8122C>T)	Stop codon [n]	51	
pkd131	I	Pregnancy termination	p.Thr257Pro (c.769A>C)	Missense [n]	11	PP
			p.Met1? (c.2T>C)	Missense [n]	2	
pkd133	I	6 m	p.Ile2957Thr (c.8870T>C)	Missense	57	Y2
			c.3365-3C>T	Splicing [n]	IVS29	
pkd134	I	3 m	p.Gly952Arg (c.2854G>A)	Missense	27	Y2
			p.Pro1628Ala (c.4882C>G)	Missense	32	
pkd135	I	16 y	c.6034_6039delACACTC (p.Thr2012_Leu2013del)	In frame [n]	37	Y2
			p.Asp3213Ala (c.9638A>C)	Missense [n]	58	
pkd136	I	9 m	p.Arg1624Trp (c.4870C>T)	Missense	32	Y2
			p.Gln2708* (c.8122C>T)	Stop codon [n]	51	
pkd137	I	33 y	p.Asn1779Ser (c.5336A>G)	Missense [n]	33	NA
			p.Gly1994Ala (c.5981G>C)	Missense [n]	37	
pkd138	I	3 y	p.Asn1779Ser (c.5336A>G)	Missense [n]	33	Y2
			p.Asn1779Ser (c.5336A>G)	Missense [n]	33	
pkd139	I	9 m	p.Tyr2376* (c.7128C>A)	Stop codon [n]	45	Y1
			p.Ile3167Thr (c.9500T>C)	Missense [n]	58	
pkd145	I	7 m	c.10659delT (p.Ile3553Metfs*15)	Frameshift [n]	61	Y2
			p.Ile222Val (c.664A>G)	Missense	9	
pkd152	I	Perinatal (alive)	p.Ile2957Thr (c.8870T>C)	Missense	57	Y2
			p.Gly952Arg (c.2854G>A)	Missense	27	
pkd154	I	4 y	p.Gln1473* (c.4417C>T)	Stop codon [n]	32	Y2
			p.Ser1690Pro (c.5068T>C)	Missense [n]	32	
pkd156	I	Perinatal (alive)	p.Gln1473* (c.4417C>T)	Stop codon [n]	32	Y2
			c.6122-?_6490+?del (p.Ser2042_Gly2164del)	Large deletion [n]	38–39	
pkd157	I	11 m	p.Arg19Cys (c.55C>T)	Missense [n]	3	Y2
			p.Glu2158* (c.6472G>T)	Stop codon [n]	39	
pkd159	I	25 y	p.Cys1431Tyr (c.4292G>A)	Missense	32	Y2
			c.5895dupA (p.Leu1966Thrfs*4)	Frameshift	36	
pkd163	I	13 y	c.5895dupA (p.Leu1966Thrfs*4)	Frameshift	36	Y2
			p.Val3036Gly (c.9107T>G)	Missense	58	
pkd164	I	Pregnancy termination	c.5895dupA (p.Leu1966Thrfs*4)	Frameshift	36	Y2
			p.Ile3051Thr (c.9152T>C)	Missense	58	
pkd177	I	Pregnancy termination	c.1388delC (p.Pro463Glnfs*5)	Frameshift	16	PP
			p.Arg3107* (c.9319C>T)	Stop codon	58	
pkd179	I	17 y	p.Arg1624Trp (c.4870C>T)	Missense	32	Y1
			c.823dupA (p.Thr275Asnfs*11)	Frameshift [n]	12	
pkd180	I	40 y	p.Asp1944Asn (c.5981G>C)	Missense [n]	36	Y1
			c.1603-1delG	Splicing [n]	IVS17	
pkd183	I	36 y	c.5761_5765delTCTTT (p.Ser1921Thrfs*15)	Frameshift	36	Y2
			p.Ile2427Thr (c.7280T>C)	Missense	46	
pkd189	I	5 m	p.Arg124* (c.370C>T)	Stop codon	5	NA
			p.Thr713Ala (c.2137A>G)/p.Leu2244His (c.6731T>A)	Missense [n]/missense [n]	21/41	
pkd190	I	18 y	p.Gln1155* (c.3463C>T)	Stop codon [n]	30	Y2
			c.6122-?_6808+?del (p.Ser2042_Gly2270del)	Large deletion [n]	38–41	

Table 2 (Continued)

Patient	*Ethnic origin*	*Age at enrollment/death*	*PKHD1 mutation*	*Type of mutation*	*Exon*	*Parent analysis*
pkd194	M	Neonatal demise	p.Glu218Lys (c.652G>A)	Missense	9	Y2
			p.Thr1781Ser (c.5341A>T)	Missense [n]	33	
pkd195	I	9 m	p.Arg1775* (c.5323C>T)	Stop codon	33	Y2
			p.Asn901Thr (c.2702A>C)	Missense	25	
pkd199	I	Neonatal demise	c.8554+1G>A	Splicing [n]	IVS54	NA
			p.Leu3706Arg (c.11117T>G)	Missense [n]	61	
pkd202	I	Neonatal demise	p.Thr36Met (c.107C>T)	Missense	5	PP
			p.Leu3713* (c.11138T>A)	Stop codon [n]	61	
pkd207	M	4 y	p.Arg1624Trp (c.4870C>T)	Missense	32	Y2
			p.Arg1624Trp (c.4870C>T)	Missense	32	
pkd208	I	2 y	p.Arg1624Trp (c.4870C>T)	Missense	32	NA
			p.Arg2671* (c.8011C>T)	Stop codon	50	
pkd210	I	17 y	p.Ile2957Thr (c.8870T>C)	Missense	57	Y2
			p.Pro1710Ser (c.5128C>T)	Missense [n]	32	
pkd212	I	12 m	p.Arg1775* (c.5323C>T)	Stop codon	33	Y2
			p.His2655Arg (c.7964A>G)	Missense [n]	50	
pkd213	M	Neonatal demise	p.Arg124* (c.370C>T)	Stop codon	5	NA
			p.Arg124* (c.370C>T)	Stop codon	5	
pkd214	I	7 y	p.Thr36Met (c.107C>T)	Missense	6	NA
			p.Ala3224Pro (c.9670G>C)	Missense [n]	58	
pkd216	A	Neonatal demise	p.Thr36Met (c.107C>T)	Missense	7	Y2
			p.Arg760His (c.2279G>A)	Missense	22	
pkd219	M	13 y	c.2339_2340delGA (p.Arg780Thrfs*18)	Frameshift [n]	23	NA
			p.Ser270Arg (c.808A>C)	Missense [n]	12	
pkd220	I	9 y	p.Gln57* (c.169C>T)	Stop codon [n]	4	NA
			c.2407+4A>G	Splicing [n]	IVS23	
pkd221	P	Pregnancy termination	p.Pro1940Ser (c.5818C>T)	Missense [n]	36	PP
			p.Pro1940Ser (c.5818C>T)	Missense [n]	36	
pkd245	I	19 y	c.707+1G>A	Splicing [n]	IVS10	Y1
			p.His2655Arg (c.7964A>G)	Missense [n]	50	
pkd4	I	3 y	p.His3124Tyr (c.9370C>T)	Missense	58	NA
			(?)	–	–	
pkd6	I	3 y	p.Arg1775* (c.5323C>T)	Stop codon	33	NA
			(?)	–	–	
pkd13	I	10 y	p.Arg1775* (c.5323C>T)	Stop codon	33	NA
			(?)	–	–	
pkd19	I	19 y	p.Ile222Val (c.664A>G)	Missense	9	NA
			(?)	–	–	
pkd27	I	Neonatal demise	p.Arg1775* (c.5323C>T)	Stop codon	33	NA
			(?)	–	–	
pkd36	I	Neonatal demise	p.Met1? (c.1A>G)	Missense [n]	2	PP
			(?)	–	–	
pkd43	I	Neonatal demise	p.Thr2641Ala (c.7921A>G)	Missense	50	Y2
			(?)	–	–	
pkd46	I	Neonatal demise	p.Ser3018Phe (c.9053C>T)	Missense	58	Y2
			(?)	–	–	
pkd48	I	Pregnancy termination	p.Arg3772* (c.11314C>T)	Stop codon	63	Y2
			(?)	–	–	
pkd53	I	7 m	p.Arg3556Cys (c.10666C>T)	Missense	61	NA
			(?)	–	–	
pkd55	I	10 y	p.Ser1400Leu (c.4199C>T)	Missense [n]	32	Y2
			(?)	–	–	
pkd60	I	13 y	p.Thr2235Arg (c.6704C>G)	Missense [n]	41	Y2
			(?)	–	–	
pkd68	I	Neonatal demise	p.Arg760His (c.2279G>A)	Missense	22	PP
			(?)	–	–	
pkd70	I	4 y	p.Arg1624Trp (c.4870C>T)	Missense	32	Y2
			(?)	–	–	
pkd71	I	12 m	p.Gly223Ser (c.667G>A)	Missense	9	Y2
			(?)	–	–	

Table 2 (Continued)

Patient	*Ethnic origin*	*Age at enrollment/death*	*PKHD1 mutation*	*Type of mutation*	*Exon*	*Parent analysis*
pkd76	I	Perinatal (alive)	p.Asn2300Lys (c.6900C>G)	Missense [n]	43	Y2
			(?)	–	–	
pkd77	I	Perinatal (alive)	p.Pro1628Ala (c.4882C>G)	Missense	32	Y1
			(?)	–	–	
pkd83	I	4 y	p.Ala3072Val (c.9215C>T)	Missense	58	Y2
			(?)	–	–	
pkd84	I	Neonatal demise	p.Tyr1838Cys (c.5513A>G)	Missense	34	PP
			(?)	–	–	
pkd89	I	Neonatal demise	p.Tyr255Asn (c.763T>A)	Missense [n]	11	Y2
			(?)	–	–	
pkd98	I	12 m	p.Gly1994Ala (c.5981G>C)	Missense [n]	37	Y2
			(–)	–	–	
pkd100	I	2 y	c.6122-?_6808+?del (p.Ser2042_Gly2270del)	Large deletion [n]	38–41	NA
			(–)	–	–	
pkd106	I	2 y	c.5895dupA (p.Leu1966Thrfs*4)	Frameshift	36	NA
			(–)	–	–	
pkd113	I	Neonatal demise	p.Gly466Glu (c.1397G>A)	Missense	16	PP
			(–)	–	–	
pkd122	A	Neonatal demise	c.5895dupA (p.Leu1966Thrfs*4)	Frameshift	36	Y2
			(–)	–	–	
pkd124	I	2 y	p.Glu1448Gly (c.4343A>G)	Missense #	32	Y1
			(–)	–	–	
pkd127	I	13 y	p.Arg1624Trp (c.4870C>T)	Missense	32	Y1
			(–)	–	–	
pkd132	I	5 y	p.Ile222Val (c.664A>G)	Missense	9	Y2
			(–)	–	–	
pkd140	I	Perinatal (alive)	p.Gly2041Ser (c.6121G>A)	Missense [n]	37	NA
			(–)	–	–	
pkd141	M	Perinatal (alive)	p.Trp1229Ser (c.3686G>C)	Missense #	32	NA
			(–)	–	–	
pkd143	I	Neonatal demise	p.Val3036Gly (c.9107T>G)	Missense	58	Y2
			(–)	–	–	
pkd148	I	3 m	c.9523_9524delinsG (p.Asn3175Valfs*8)	Frameshift [n]	58	Y2
			(–)	–	–	
pkd151	I	20 y	p.Ile222Val (c.664A>G)	Missense	9	NA
			(–)	–	–	
pkd160	I	36 y	c.1197delC (p.Leu400Cysfs*13)	Frameshift [n]	15	NA
			(–)	–	–	
pkd172	I	Pregnancy termination	p.Asn3384Lys (c.10152C>A)	Missense [n]	60	PP
			(–)	–	–	
pkd173	I	12 m	p.Asp1944Asn (c.5981G>C)	Missense [n]	36	NA
			(–)	–	–	
pkd176	I	3 m	p.Cys1431Tyr (c.4292G>A)	Missense	32	Y2
			(–)	–	–	
pkd178	I	54 y	p.Pro2356Leu (c.7067C>T)	Missense	44	NA
			(–)	–	–	
pkd198	I	3 y	p.Asn901Thr (c.2702A>C)	Missense	25	Y2
			(–)	–	–	
pkd204	I	9 y	p.Cys1431Tyr (c.4292G>A)	Missense	32	NA
			(–)	–	–	
pkd205	C	Neonatal demise	p.Pro755Leu (c.2264C>T)	Missense	22	NA
			(–)	–	–	
pkd238	I	6 m	p.Arg2955Gln (c.8864G>A)	Missense [n]	57	Y2
			(–)	–	–	
pkd240	I	Neonatal demise	p.Gly3055Cys (c.9163G>T)	Missense [n]	58	Y2
			(–)	–	–	
pkd94	I	12 m	(–)	–	–	NA
			(–)	–	–	
pkd103	I	9 y	(–)	–	–	NA
			(–)	–	–	

Table 2 (Continued)

Patient	*Ethnic origin*	*Age at enrollment/death*	*PKHD1 mutation*	*Type of mutation*	*Exon*	*Parent analysis*
pkd104	I	4 m	(−)	–	–	NA
			(−)	–	–	
pkd112	I	8 y	(−)	–	–	NA
			(−)	–	–	
pkd129	A	48 y	(−)	–	–	NA
			(−)	–	–	
pkd149	I	12 m	(−)	–	–	NA
			(−)	–	–	
pkd161	M	15 y	(−)	–	–	NA
			(−)	–	–	
pkd162	I	Perinatal (alive)	(−)	–	–	NA
			(−)	–	–	
pkd168	I	10 y	(−)	–	–	NA
			(−)	–	–	
pkd170	I	4 m	(−)	–	–	NA
			(−)	–	–	
pkd171	I	Perinatal (alive)	(−)	–	–	NA
			(−)	–	–	
pkd197	I	7 m	(−)	–	–	NA
			(−)	–	–	

Abbreviations: ARPKD, autosomal recessive polycystic kidney disease; m, month; *PKHD1*, *polycystic kidney and hepatic disease 1*; y, year. Column 2: A, Albania; C, China; I, Italy; M, Morocco; P, Pakistan. Column 4: (?), absence of mutation in patients not completely analyzed; (−), absence of mutations in patients completely analyzed. Column 5: #, possible mutations (see test); [n], novel mutation identified in this work. Column 7: NA, not available; PP, only patients' parents available for analysis; Y1, one parent analyzed; Y2, both parents analyzed.

RESULTS

Mutation analysis

A total of 130 unrelated subjects underwent genetic testing for ARPKD, of whom 117 (90%) were Italian; the remaining patients were from Morocco, Albania, China and Pakistan (6, 4, 2 and 1, respectively).

Mutation analysis of *PKHD1* gene was carried out by direct sequencing, applying the algorithm described by Bergmann *et al.*[36] and MLPA technique. We performed an exhaustive analysis in 110 out of 130 individuals, whereas in the other 20 subjects, all of Italian origin, a partial screening was carried out because of insufficient and not more available DNA sample. For each of these 20 subjects, a number of *PKHD1* exons ranging from 12 to 45 were analyzed by sequencing.

Taking into account the 110 patients who were completely analyzed (97 Italian and 13 from different countries), we identified 173 mutations on a total of 220 affected alleles, with a detection rate of 78.6%. Thus, in 98 families (89.1%) at least one mutation was identified. Any mutation was identified on 12 probands (Table 1). Sixty-five patients were compound heterozygous and 10 patients were homozygous. After the segregation analysis performed on the parents, in 54 out of 75 probands carrying 2 mutations, we were able to prove that the mutant alleles reside on separate chromosomes (Table 2).

On the other hand, considering the total number ($n = 130$) of analyzed individuals, 107 different types of mutations were detected in 193 mutated alleles. Out of 107 mutations, 45 had been previously described and reported in the literature: 12 truncating mutations, 1 splice site mutation and 30 missense mutations that are considered pathogenic in the public ARPKD/*PKHD1* database. Two additional variants, p.Glu1448Gly and p.Trp1229Ser with a very low frequency (minor allele frequency = 0.002 and 0.003 in ExAC database, respectively), were considered as possible mutations even if they are not clearly characterized as pathogenic on this locus-specific database.

In all, 62 were novel mutations (Tables 2 and 3); 24 were truncating mutations considered definitely pathogenic: 11 nonsense, 6 frameshift, 2 multiexon deletions detected by MLPA and 5 splice site mutations; 2 were in-frame deletions. Two splice site mutations affected nucleotide at noncanonical position of the donor and acceptor splice site, in intron 23 and intron 29, respectively. The c.2407+4A>G variant, identified in patient pkd220, was predicted to significantly alter the wild-type donor splice site according to the HSF3.0 resource (wild-type site score: 81.15; mutant site score: 72.81; variation: −10.28%). On the contrary, the c.3365-3C>T change was not predicted as dangerous variant; however, its finding in two unrelated patients pkd82 and pkd133, the segregation in their families and the absence in the public database makes it a possible pathogenic mutation rather than a neutral variant. Thirty-four were missense variants classified as likely pathogenic mutations according to the segregation analysis and *in silico* evaluation (Table 3). Figure 1 shows the map of amino acid variants along the PKHD1 protein schematic structure. In two patients, three variants were identified: the nonsense p.Arg124* mutation and the novel missense p.Thr713Ala and p.Leu2244His changes were identified in patient pkd189; the novel missense mutations p.Asp2528Asn and p.Ala3224Pro, the latter at homozygous state, were identified in patient pkd30. Unfortunately, parents of pkd189 were not available and thus it was not possible to verify the segregation pattern. The p.Ala3224Pro variant was found in homozygosity in pkd30 patient and in compound heterozygosity in pkd31 and pkd214 patients, with all these subjects belonging to the same geographical origin of North Italy. In all, 69 subjects underwent MLPA analysis: 25 carried a single variant, 13 without mutations and 31 carried novel missense mutations in homozygosity or in compound heterozygosity. These latter were included in order to exclude the presence of deleterious mutations different from the novel missense found. In total, four alleles were characterized. A heterozygous deletion of the exons 38 and 39 (c.6122-?_6490+?del) was identified in patients pkd111 and pkd156 and confirmed in the available parents.

Table 3 Description of the putative effect of missense *PKHD1* mutations

Amino acid substitution	*Genomic variation (hg19)*	*rsID dbSNP144*	*1000Gp3AF*	*ExAC0.3AF*	*SIFT*	*Polyphen2 (HDIV)*	*Polyphen2 (HVAR)*	*LRT*	*Mutation Taster*	*Mutation Assessor*	*FATHMM*	*PROVEAN*	*CADD (phred-scaled)*	*Classification resume*	*GERP+ + RS*	*phyloP7way (vertebrate)*
p.Thr257Pro	chr6:g.51934264T>G	—	—	—	D	D	D	D	D	M	D	D	26.3	9D/0T	5.28	0.991
p.Gly2041Ser	chr6:g.51798908C>T	rs199589074	5.99E−04	9.06E−05	D	D	D	D	D	M	D	D	20.4	9D/0T	5.5	0.871
p.Asp2528Asn	chr6:g.51732812C>T	rs142487082	—	4.12E−05	D	D	D	D	D	M	D	D	31	9D/0T	5.67	0.871
p.His2655Arg	chr6:g.51712716T>C	—	—	—	D	D	D;D;P	D	D	M	D	D	24.6	9D/0T	5.47	0.991
p.Arg2955Gln	chr6:g.51618085C>T	—	—	—	D	D	D	D	D	M	D	D	34	9D/0T	5.75	0.871
p.Gly3055Cys	chr6:g.51613251C>A	—	—	—	D	D	D	D	D	M	D	D	35	9D/0T	5.86	0.871
p.Asp3213Ala	chr6:g.51612776T>G	—	—	—	D	D	D	D	D	M	D	D	27.7	9D/0T	5.75	0.991
p.Leu3706Pro	chr6:g.51523807A>G	—	—	—	D	D	D	D	D	M	D	D	26.6	9D/0T	5.91	1.062
p.Tyr255Asn	chr6:g.51934270A>T	—	—	—	D	D	D	N	D	M	D	D	27.5	8D/1T	5.28	1.062
p.Ser270Arg	chr6:g.51930846T>G	—	—	—	D	D	D	D	D	M	T	D	22.2	8D/1T	5.95	0.991
p.Thr713Ala	chr6:g.51917877T>C	—	—	—	D	D	P	D	D	M	D	D	24.6	8D/1T	5.63	0.991
p.Pro1940Ser	chr6:g.51824758G>A	—	—	—	D	D	D	D	D	M	D	D	17.59	8D/1T	5.69	0.917
p.Asp1944Asn	chr6:g.51824746C>T	—	—	1.65E−05	D	D	D	D	D	M	D	N	20.2	8D/1T	5.69	0.871
p.Gly1994Ala	chr6:g.51799048C>G	—	—	—	D	D	D	D	D	M	D	D	18.51	8D/1T	5.5	0.871
p.Asn2300Lys	chr6:g.51768491G>C	—	—	—	D	D	D	D	N	M	D	D	22.9	8D/1T	−0.504	0.045
p.Pro1710Leu	chr6:g.51889479G>A	—	—	4.94E−05	D	P;D	P	N	D	M	D	D	24.1	7D/2T	4.68	0.917
p.Ala3224Pro	chr6:g.51612744C>G	—	—	—	T	D	D	D	D	M	D	N	24.4	7D/2T	5.75	0.871
p.Cys1431Tyr	chr6:g.51890316C>T	—	—	3.30E−05	D	P;D	B;P	N	N	M	D	D	24.2	6D/3T	5.01	0.824
p.Ser1690Pro	chr6:g.51889540A>G	—	—	—	D	D	D	N	N	M	D	D	19.61	6D/3T	4.26	1.062
p.Leu2244His	chr6:g.51771090A>T	—	—	8.24E−06	D	D	D	N	D	M	T	D	27.7	6D/3T	5.25	1.062
p.Ile3167Thr	chr6:g.51612914A>G	—	—	8.24E−06	D	P	P;B;B	N	D	M	D	D	25	6D/3T	5.86	1.062
p.Asn3384Lys	chr6:g.51609187G>T	—	—	8.24E−06	D	D	P	N	D	L	D	D;N	21.7	6D/3T	0.0988	−0.715
p.Asn1779Ser	chr6:g.51887643T>C	—	—	—	D	P;D	B;D	D	N	L	D	D	16.9	5D/4T	5.7	0.991
p.Ser1929Pro	chr6:g.51824791A>G	—	—	—	D	D	D	N	N	N	D	D	16.96	5D/4T	5.54	1.062
p.Val2921Gly	chr6:g.51619617A>C	—	—	—	D	P	B	N	N	M	D	D	22.7	5D/4T	5.63	1.003
p.Ser1400Leu	chr6:g.51890409G>A	rs191201723	2.00E−04	8.24E−06	D	P	B	N	N	M	T	D	23.8	4D/5T	5.01	0.866
p.Thr1781Ser	chr6:g.51887638T>A	—	—	—	T	P;D	P;D	N	N	M	D	N	16.59	4D/5T	4.5	0.991
p.Met1?	chr6:g.51949731T>C	rs376987651	—	2.47E−05	D	B	B	D	N	—	D	N	22.6	4D/4T	5	0.991
p.Arg723His	chr6:g.51915066C>T	rs150597050	—	8.24E−06	T	B	B	N	N	L	D	D	23.2	3D/6T	3.76	−0.032
p.Arg19Cys	chr6:g.51948051G>A	—	—	2.47E−05	D	P	B	N	N	N	D	N	17.25	2D/7T	−2.88	−0.463
p.Glu1202Lys	chr6:g.51892651C>T	rs554956088	2.00E−04	4.12E−05	T	D	P;B	N	N	L	T	N	22.7	2D/7T	0.787	0.871
p.Ala1758Gly	chr6:g.51887706G>C	—	—	—	T	B	B	N	N	L	D	N	15.44	1D/8T	3.95	−0.044
p.Thr2235Arg	chr6:g.51771117G>C	—	—	—	T	P	B	N	N	L	D	N	18.98	1D/8T	4.38	0.917
p.Ser3524Phe	chr6:g.51524353G>A	—	—	—	T	B	B	N	N	N	D	N	11.91	1D/8T	4.84	0.034

Abbreviation: *PKHD1*, *polycystic kidney and hepatic disease 1*. First column: protein change. Column 3: accession number for known variants; Column 4: allele frequency within 1000Genomes samples; Column 5: allele frequency within the Exome Aggregation Consortium (ExAC) exome samples. Columns 6–13: categorical pathogenicity predictions for the different methods: D, deleterious/damaging; B, benign; N, neutral; P', probably damaging; M, medium impact; L, low impact; T, tolerated; '—', unavailable prediction; P;D, multiple predictions if protein is present with multiple entries. Column 14: phred-scaled CADD score; scores >20 are supposed to be damaging. Column 15: resume of variant pathogenicity predictions: 'T' means Tolerated and it has been attributed for B, N, P, P;B, B;D, B;P, L, T and P;B;B categories (columns 6–13); 'D' means Damaging and it is relative to D, M, P;D, D;D;P and D;N categories (columns 6–13); 'D' has also been attributed to variants with CADD scores >20. Column 16: GERP++ scores, with higher values indicating an higher degree of conservation. Values range from −12.3 to 6.7. Column 17: phyloP score based on the alignment of 7 vertebrate genome (including human); positive values indicates a more evolutionary constrained site, whereas negative values are assigned to more relaxed sites. Score range from −5.172 to 1.062 as for the queried dbNSFP version.

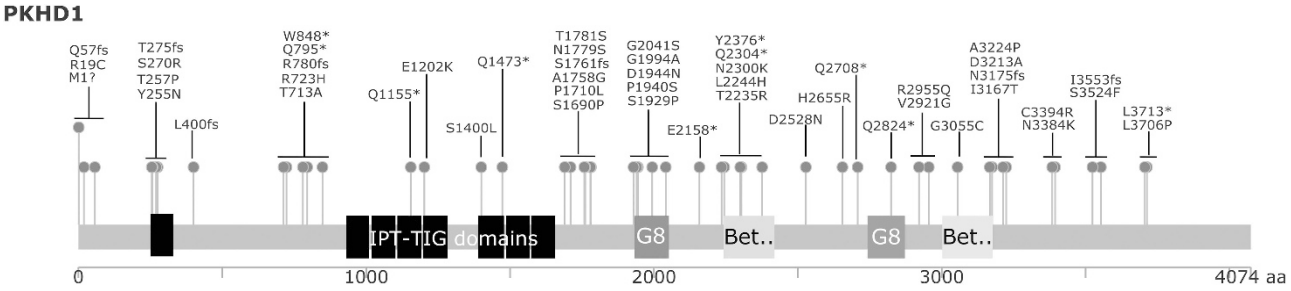

Figure 1 Map of amino acid variants along the PKHD1 (polycystic kidney and hepatic disease 1) protein schematic structure. Novel missense, frameshift and nonsense variants were considered and mapped on PKHD1 protein 'lollipop plot', by using the MutationMapper v1.01. webtool (www.cbioportal.org). Plot was downloaded as 'scalable vector graphics' file format and refined appropriately. Gray-scale blocks represent IPT/TIG, G8 and β-helix domains, respectively.

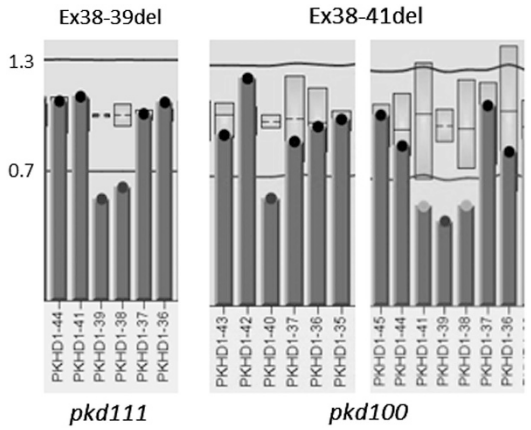

Figure 2 Characterization of two PKHD1 (polycystic kidney and hepatic disease 1) exonic deletions. Bar chart of the multiple ligation probe amplification (MLPA) analysis performed on patients pkd111 and pkd100 showing the deletion of exons 38–39 and 38–41 of the *PKHD1* gene, respectively. The normal range of the rapport of the dosage is defined by 0.7 and 1.3 values.

A second large deletion encompassing the exons 38–41 (c.6122-?_6808 +?del) was found in patients pkd100 and pkd190 (Figure 2). At the protein level, the deletion of exons 38–39 (p.Ser2042_Gly2164del) partially affects the G8 domain[50] (amino acid positions: 1933–2052), whereas the deletion of exons 38–41 (p.Ser2042_Gly2270del) affects the region between the G8 and the β-helix secondary motif (residues 1933–2052 and 2242–2419, respectively).

In addition, 10 subjects showed an apparent deletion of the exon 33 or exon 57 because of the interference of mutations p.Arg1775*, p.Asn1779Ser, p.Asn1781Ser and p.Ile2957Thr with MLPA reaction.

DISCUSSION

Here we present one of the largest studies on genetic mutations in ARPKD and representing the first screening of the *PKHD1* gene in a large cohort of Italian patients. In fact, only 21 subjects of Italian origin were studied and reported up to now in the public ARPKD/*PKHD1* database.

In order to identify causative mutations in our cohort of 130 probands, we applied the algorithm defined by Bergman *et al.*[36] to sequence all coding regions of the gene. Moreover, to search for exon deletions/duplications, MLPA analysis was carried out in a subset of subjects who resulted negative or carried only one mutation after sequencing analysis.

As only 13 out of 130 probands (10%) analyzed in this study were not Italian, we believe these results can be considered representative of the Italian population.

In the completely analyzed 110 ARPKD probands, 173 out of 220 expected mutated alleles were characterized, achieving a detection rate of 78.6%, in line with whole-coding sequence screenings of the *PKHD1* gene in cohorts of similar size.[13,14] Failure to detect mutations in 21.4% of chromosomes may have several explanations. The missing mutations in heterozygous subjects and in 12 patients without identified mutations could be located in deep intronic or other regulatory regions distant from the splice donor and acceptor sites that have not been screened so far; alternatively, causative mutations could reside in other genes such as *HNF1B*, *PKD1* and *PKD2*, not analyzed in the present study.

If we consider all 130 probands, a total of 107 different types of mutations have been detected accounting for a total of 193 characterized alleles. Out of 107 mutations, 45 (42.1%) had been previously described and reported in the literature, whereas 62 (57.9%) were novel. The most frequent mutations identified in our cohort were: c.5895dupA (p.Leu1966Thrfs*4) and p.Arg1624Trp (4.7%), p.Thr36Met and p.Arg1775* (3.6%), p.Ile222Val (3.1%), among the known mutations and the p.Asn1779Ser (2.6%) and the p.Gln2708* (2.1%) among the novel. The use of MLPA technique to search for *PKHD1* deletions or duplications allowed us to identify 4 alleles (2%) and it has proved to be indispensable to reach the conclusive molecular diagnosis in 3 patients. The absence of a founder mutation reveals the high degree of the genetic heterogeneity of the Italian population. Indeed, taking into account the families carrying two mutated alleles, only 5 out of 67 (7.4%) of the Italian probands showed homozygosity, not due to consanguineous marriages, compared with the 5 out of 8 (62%) of the non-Italian probands (Table 1). Before our work, Krall *et al.*[29] applied the algorithm defined by Bergmann *et al.*[36] to analyze 50 individuals affected by ARPKD and proposed a sequencing strategy in order to facilitate genetic testing for Hispanic populations. They found that only 21 exons were sufficient to identify 86% of the expected mutated alleles. In our Italian population mutations have been detected in 45 of the 67 *PKHD1* exons with more than half of them concentrated in only 7 exons (32, 36, 58, 33, 61, 3 and 9), thus providing a 51% of chance to find at least one mutation.

In conclusion, our report expands the spectrum of *PKHD1* mutations and confirms the allelic heterogeneity of this disorder. The studied population represents the largest Italian ARPKD cohort reported to date. Nevertheless, we think the present could be one of the last screening carried out by Sanger sequencing in such large cohort of patients. In fact, genes of large size and presenting allelic

heterogeneity such as *PKHD1* make the Sanger-based sequencing of single amplicons labor intensive and expensive and fully justify the use of next-generation sequencing (NGS) technologies. In addition, the NGS approach makes it possible to evaluate several genes of interest, (that is, *HNF1B*, *PKD1* and *PKD2* in the case of ARPKD) in a single test. In the past years, some *PKHD1* mutations were identified by exome sequencing.[23,30,31] Recently, Tavira *et al.*[51] estimated that the cost of the NGS of the 30 samples, plus Sanger sequencing of PCR-fragments to assign the identified mutations, was ~3.5-fold lower than the Sanger sequencing of all the *PKHD1* amplicons on an ABI3130xl sequencer. Besides, Sanger sequencing-based molecular screening would have required several weeks, whereas the NGS project was completed in only 2 weeks. Exome sequencing and targeted resequencing are thus attractive tools that can improve molecular diagnosis in ARPKD and many other genetic diseases.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We gratefully acknowledge the patients and the referring physicians who participated in this study. We thank Matteo Granatiero for his technical support. This work is dedicated to the memory of Dr Leopoldo Zelante who died in December (1948–2015).

1 Zerres, K., Rudnik-Schoneborn, S., Steinkamm, C., Becker, J. & Mucher, G. Autosomal recessive polycystic kidney disease. *J. Mol. Med.* **76**, 303–309 (1998).
2 Guay-Woodford, L. M. & Desmond, R. A. Autosomal recessive polycystic kidney disease: the clinical experience in North America. *Pediatrics* **111**, 1072–1080 (2003).
3 Fonck, C., Chauveau, D., Gagnadoux, M. F., Pirson, Y. & Grunfeld, J. P. Autosomal recessive polycystic kidney disease in adulthood. *Nephrol. Dial. Transplant.* **16**, 1648–1652 (2001).
4 Blyth, H. & Ockenden, B. G. Polycystic disease of kidney and liver presenting in childhood. *J. Med. Genet.* **8**, 257–284 (1971).
5 Guay-Woodford, L. M., Muecher, G., Hopkins, S. D., Avner, E. D., Germino, G. G., Guillot, A. P. *et al.* 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 Ward, C. J., Hogan, M. C., Rossetti, S., Walker, D., Sneddon, T., Wang, X. *et al.* The gene mutated in autosomal recessive polycystic kidney disease encodes a largE., receptor-like protein. *Nat. Genet.* **30**, 259–269 (2002).
7 Onuchic, L. F., Furu, L., Nagasawa, Y., Hou, X., Eggermann, T., Ren, Z. *et al. 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).
8 Bergmann, C., Senderek, J., Sedlacek, B., Pegiazoglou, I., Puglia, P., Eggermann, T. *et al.* Spectrum of mutations in the gene for autosomal recessive polycystic kidney disease (ARPKD/*PKHD1*). *J. Am. Soc. Nephrol.* **14**, 76–89 (2003).
9 Lambie, L., Amin, R., Essop, F., Cnaan, A., Krause, A. & Guay-Woodford, L. M. Clinical and genetic characterization of a founder *PKHD1* mutation in Afrikaners with ARPKD. *Pediatr. Nephrol.* **30**, 273–279 (2015).
10 Furu, L., Onuchic, L. F., Gharavi, A., Hou, X., Esquivel, E. L., Nagasawa, Y. *et al.* Milder presentation of recessive polycystic kidney disease requires presence of amino acid substitution mutations. *J. Am. Soc. Nephrol.* **14**, 2004–2014 (2003).
11 Rossetti, S., Torra, R., Coto, E., Consugar, M., Kubly, V., Malaga, S. *et al.* A complete mutation screen of *PKHD1* in autosomal-recessive polycystic kidney disease (ARPKD) pedigrees. *Kidney Int.* **64**, 391–403 (2003).
12 Bergmann, C., Senderek, J., Küpper, F., Schneider, F., Dornia, C., Windelen, E. *et al. PKHD1* mutations in autosomal recessive polycystic kidney disease (ARPKD). *Hum. Mutat.* **23**, 453–463 (2004).
13 Bergmann, C., Senderek, J., Schneider, F., Dornia, C., Kupper, F., Eggermann, T. *et al. PKHD1* mutations in families requesting prenatal diagnosis for autosomal recessive polycystic kidney disease (ARPKD). *Hum. Mutat.* **23**, 487–495 (2004).
14 Losekoot, M., Haarloo, C., Ruivenkamp, C., White, S. J., Breuning, M. H. & Peters, D. J. Analysis of missense variants in the *PKHD1*-gene in patients with autosomal recessive polycystic kidney disease (ARPKD). *Hum. Genet.* **118**, 185–206 (2005).
15 Bergmann, C., Senderek, J., Windelen, E., Kupper, F., Middeldorf, I., Schneider, F. *et al.* Clinical consequences of *PKHD1* mutations in 164 patients with autosomal-recessive polycystic kidney disease (ARPKD). *Kidney Int.* **67**, 829–848 (2005).
16 Bergmann, C., Kupper, F., Schmitt, C. P., Vester, U., Neuhaus, T. J., Senderek, J. *et al.* Multi-exon deletions of the *PKHD1* gene cause autosomal recessive polycystic kidney disease (ARPKD). *J. Med. Genet.* **42**, e63 (2005).
17 Adeva, M., El-Youssef, M., Rossetti, S., Kamath, P. S., Kubly, V., Consugar, M. B. *et al.* Clinical and molecular characterization defines a broadened spectrum of autosomal recessive polycystic kidney disease (ARPKD). *Medicine (Baltimore)* **85**, 1–21 (2006).
18 Sharp, A. M., Messiaen, L. M., Page, G., Antignac, C., Gubler, M. C., Onuchic, L. F. *et al.* Comprehensive genomic analysis of *PKHD1* mutations in ARPKD cohorts. *J. Med. Genet.* **42**, 336–349 (2005).
19 Michel-Calemard, L., Dijoud, F., Till, M., Lambert, J. C., Vercherat, M., Tardy, V. *et al.* Pseudoexon activation in the *PKHD1* gene: a French founder intronic mutation IVS46 +653A>G causing severe autosomal recessive polycystic kidney disease. *Clin. Genet.* **75**, 203–206 (2009).
20 Gunay-Aygun, M., Tuchman, M., Font-Montgomery, E., Lukose, L., Edwards, H., Garcia, A. *et al. PKHD1* sequence variations in 78 children and adults with autosomal recessive polycystic kidney disease and congenital hepatic fibrosis. *Mol. Genet. Metab.* **99**, 160–173 (2010).
21 Denamur, E., Delezoide, A. L., Alberti, C., Bourillon, A., Gubler, M. C., Bouvier, R. *et al.* Genotype-phenotype correlations in fetuses and neonates with autosomal recessive polycystic kidney disease. *Kidney Int.* **77**, 350–358 (2010).
22 Zvereff, V., Yao, S., Ramsey, J., Mikhail, F. M., Vijzelaar, R. & Messiaen, L. Identification of *PKHD1* multiexon deletions using multiplex ligation-dependent probe amplification and quantitative polymerase chain reaction. *Genet. Test. Mol. Biomarkers* **14**, 505–510 (2010).
23 Zhang, D., Lu, L., Yang, H. B., Li, M., Sun, H., Zeng, Z. P. *et al.* Exome sequencing identifies compound heterozygous *PKHD1* mutations as a cause of autosomal recessive polycystic kidney disease. *Chin. Med. J.* **125**, 2482–2486 (2012).
24 Zhou, X. H., Hui, Z. Y. & Li, Y. Clinical and pathological features of a neonate with autosomal recessive polycystic kidney disease caused by a nonsense *PKHD1* mutation. *World. J. Pediatr.* **9**, 76–79 (2013).
25 Xu, Y., Xiao, B., Jiang, W. T., Wang, L., Gen, H. Q., Chen, Y. W. *et al.* A novel mutation identified in *PKHD1* by targeted exome sequencing: guiding prenatal diagnosis for an ARPKD family. *Gene* **551**, 33–38 (2014).
26 Hao, X., Liu, S., Dong, Q., Zhang, H., Zhao, J. & Su, L. Whole exome sequencing identifies recessive *PKHD1* mutations in a Chinese twin family with Caroli disease. *PLoS ONE* **9**, e92661 (2014).
27 Liu, S. P., Ding, J., Wang, F., Zhang, Y. Q. & Ye, J. T. Clinical characteristics and mutation analysis of three Chinese children with autosomal recessive polycystic kidney disease. *World J. Pediatr.* **10**, 271–274 (2014).
28 Thakur, P., Speer, P. & Rajkovic, A. Novel mutation in the *PKHD1* gene diagnosed prenatally in a fetus with autosomal recessive polycystic kidney disease. *Case Rep. Genet* **2014**, 517952 (2014).
29 Krall, P., Pineda, C., Ruiz, P., Ejarque, L., Vendrell, T., Camacho, J. A. *et al.* Cost-effective *PKHD1* genetic testing for autosomal recessive polycystic kidney disease. *Pediatr. Nephrol.* **29**, 223–234 (2014).
30 Nabhan, M. M., Abdelaziz, H., Xu, Y., El Sayed, R., Santibanez-Koref, M., Soliman, N. A. *et al.* Case report: whole-exome analysis of a child with polycystic kidney disease and ventriculomegaly. *Genet. Mol. Res.* **14**, 3618–3624 (2015).
31 Herbst, S. M., Schirmer, S., Posovszky, C., Jochum, F., Rödl, T., Schroeder, J. A. *et al.* Taking the next step forward - diagnosing inherited infantile cholestatic disorders with next generation sequencing. *Mol. Cell. Probes* **29**, 291–298 (2015).
32 Miyazaki, J., Ito, M., Nishizawa, H., Kato, T., Minami, Y., Inagaki, H. *et al.* Intragenic duplication in the *PKHD1* gene in autosomal recessive polycystic kidney disease. *BMC Med. Genet.* **16**, e98 (2015).
33 Rossetti, S., Kubly, V. J., Consugar, M. B., Hopp, K., Roy, S., Horsley, S. W. *et al.* Incompletely penetrant PKD1 alleles suggest a role for gene dosage in cyst initiation in polycystic kidney disease. *Kidney Int.* **75**, 848–855 (2009).
34 Vujic, M., Heyer, C. M., Ars, E., Hopp, K., Markoff, A., Orndal, C. *et al.* Incompletely penetrant PKD1 alleles mimic the renal manifestations of ARPKD. *J. Am. Soc. Nephrol.* **21**, 1097–1102 (2010).
35 Rossetti, S., Hopp, K., Sikkink, R. A., Sundsbak, J. L., Lee, Y. K., Kubly, V. *et al.* Identification of gene mutations in autosomal dominant polycystic kidney disease through targeted resequencing. *J. Am. Soc. Nephrol.* **23**, 915–933 (2012).
36 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). *Hum. Mutat.* **25**, 225–231 (2005).
37 Liu, X., Jian, A. X., Boerwinkle, E. & Jian, X. dbNSFP: a lightweight database of human non-synonymous SNPs and their functional predictions. *Hum. Mutat.* **32**, 894–899 (2011).
38 Ng, P.C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic. Acids Res.* **31**, 3812–3814 (2003).
39 Choi, Y., Sims, G. E., Murphy, S., Miller, J. R & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* **7**, e46688 (2012).
40 Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
41 Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).
42 Schwarz, J. M., Cooper, D. N., Schuelke, M. & Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* **11**, 361–362 (2014).
43 Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**, e118 (2011).

44 Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L., Edwards, K. J. *et al*. Predicting the functional, molecular and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.* **34**, 57–65 (2013).
45 Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M & Shendure, J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
46 Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
47 Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M, Sidow, A. & Batzoglou, S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* **6**, e1001025 (2010).
48 Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L. & Smigielski, E. M. dbSNP: the NCBI database of genetic variation. *Nucl. Acids Res.* **29**, 308–311 (2001).
49 Desmet, F. O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M. & Béroud, C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, e67 (2009).
50 He, Q. Y., Liu, X. H., Li, Q., Studholme, D. J., Li, X. W. & Liang, S. P. G8: a novel domain associated with polycystic kidney disease and non-syndromic hearing loss. *Bioinformatics* **22**, 2189–2191 (2006).
51 Tavira, B., Gómez, J., Málaga, S., Santos, F., Fernández-Aracama, J., Alonso, B. *et al*. A labor and cost effective next generation sequencing of *PKHD1* in autosomal recessive polycystic kidney disease patients. *Gene* **561**, 65–169 (2015).