Am. J. Hum. Genet. 65:39–49, 1999

Identification of Mutations in the Repeated Part of the Autosomal Dominant Polycystic Kidney Disease Type 1 Gene, PKD1, by Long-Range PCR

Ruth Thomas,[1] Robert McConnell,[2] Jo Whittacker,[1] Peter Kirkpatrick,[2] John Bradley,[3] and Richard Sandford[1]

Departments of [1]Medical Genetics, [2]Neurosurgery, and [3]Medicine, Addenbrooke's Hospital, Cambridge

Summary

We have used long-range PCR to identify mutations in the duplicated part of the PKD1 gene. By means of a PKD1-specific primer in intron 1, an ~13.6-kb PCR product that includes exons 2–15 of the PKD1 gene has been used to search for mutations, by direct sequence analysis. This region contains the majority of the predicted extracellular domains of the PKD1-gene product, polycystin, including the 16 novel PKD domains that have similarity to immunoglobulin-like domains found in many cell-adhesion molecules and cell-surface receptors. Direct sequence analysis of exons encoding all the 16 PKD domains was performed on PCR products from a group of 24 unrelated patients with autosomal dominant polycystic kidney disease (ADPKD [MIM 173900]). Seven novel mutations were found in a screening of 42% of the PKD1-coding region in each patient, representing a 29% detection rate; these mutations included two deletions (one of 3 kb and the other of 28 bp), one single-base insertion, and four nucleotide substitutions (one splice site, one nonsense, and two missense). Five of these mutations would be predicted to cause a prematurely truncated protein. Two coding and 18 silent polymorphisms were also found. When, for the PKD1 gene, this method is coupled with existing mutation-detection methods, virtually the whole of this large, complex gene can now be screened for mutations.

Received February 5, 1999; accepted for publication May 13, 1999; electronically published June 9, 1999.

Address for correspondence and reprints: Dr. Richard Sandford, Department of Medical Genetics, Cambridge Institute of Medical Research, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2XY, United Kingdom. E-mail: rns13@cam.ac.uk

0002-9297/99/6501-0008$02.00

Introduction

Autosomal dominant polycystic kidney disease (ADPKD) is a common genetically heterogeneous disorder. It has an incidence of ~1:800 in all populations, with mutations in the PKD1 gene at 16p13.3 accounting for ~85% of cases and with those in the PKD2 gene at 4q21-23 accounting for the remainder (Peters and Sandkuijl 1992). A third possible locus, PKD3, remains to be identified (Paterson and Pei 1998). The phenotype produced by mutations in either the PKD1 gene or the PKD2 gene appears to be indistinguishable and consists principally of progressive renal enlargement and cyst formation leading, in the majority of affected individuals, to end-stage renal failure (ESRF) in later life. The disease associated with PKD1-gene mutations appears to be more severe, with earlier-onset renal failure (Hateboer et al. 1999). Extrarenal manifestations occur and include hepatic and pancreatic cysts and cerebral-artery aneurysms (Watson and Torres 1996). However, marked variation in the clinical presentation of ADPKD exists between and within families, suggesting that environmental and genetic factors may contribute to disease expression. This is supported by the demonstration of a genetic "two-hit" mechanism in renal- and hepatic-cyst formation. Germline and somatic inactivating mutations of the PKD1 gene have been identified in cyst-lining epithelial cells (Qian et al. 1996; Watnick et al. 1998*b*). Mutations in the PKD2 gene, a 5.5-kb single-copy transcript, mainly comprise nucleotide substitutions or small deletions/insertions (Mochizuki et al. 1996; Veldhuisen et al. 1997; Pei et al. 1998). The majority are predicted to cause premature stop codons. Since ~70% of the PKD1 gene is present as multiple highly homologous copies elsewhere on chromosome 16, the full range of mutations in this gene has yet to be established (The European Polycystic Kidney Disease Consortium 1994). The PKD1 gene comprises 46 exons spanning 52 kb of DNA, with exons 1–32 being duplicated in at least three other transcribed copies (Hughes et al. 1995). The majority of PKD1-gene mutations identified therefore lie within the remaining single-copy re-

gion of the gene (Peral et al. 1996*a,* 1996*b,* 1997; Daniells et al. 1998; The Human Gene Mutation Database Cardiff). Most are nucleotide substitutions and small deletions/insertions that lead to premature termination of the predicted protein product.

Exons 33–46, which constitute the single-copy region of the PKD1 gene, have been analyzed by conventional mutation-screening methods such as SSCP (Peral et al. 1996*b*). More recently, attempts have been made to define mutations in the duplicated part of the PKD1 gene, by techniques such as the protein-truncation test (PTT) and long-range PCR (LR-PCR). This has extended the number of exons examined, but a considerable part of the 5′ end of the PKD1 gene, including exons 1–15, has not been fully analyzed. In the first published report of mutations detected in the duplicated region, Peral et al. used the PTT and a nonisotopic RNase-cleavage assay after PKD1-specific reverse transcription–PCR (RT-PCR), to detect mutations in the duplicated area (Peral et al. 1997). Six mutations were identified in exons 22–32; three were small insertions/deletions, and three were nucleotide substitutions (one nonsense). Analysis of a similar region by LR-PCR, with one primer located in the single-copy region, has recently identified an unusual pattern of PKD1-gene mutation (Watnick et al. 1997). In two individuals, multiple clustered base-pair changes were seen in exon 23, which would be predicted to cause multiple amino acid substitutions. These mutations have arisen as a result of a gene-conversion event between the PKD1 gene and its homologues (Watnick et al. 1998*a*). The PTT has also been used to extend mutation analysis up to and including exon 15 (Roelfsema et al. 1996). Although this method will only detect translation-terminating mutations and not missense mutations, it identified six mutations in exon 15, four of which were small deletions, in a total of eight mutations in 155 individuals.

At the present time, only a third of the PKD1 gene can be completely screened for sequence variants, limiting both the accurate definition of mutational events in the PKD1 gene and their precise relationship, if any, to the disease phenotype. A technique that allows the specific detection of mutations in the remainder of the PKD1 gene is therefore required. The amplification of discontinuous fragments from the 3′ end of exon 15 to exon 46 has already been demonstrated (Peral et al. 1996*b,* 1997; Watnick et al. 1997, 1998*b*). With PTT analysis of exons up to and including exon 15, terminating mutations can also be identified (Peral et al. 1997; Roelfsema et al. 1997). However, the detection rate has been disappointingly low. To study the complete range of PKD1-gene mutations, the remaining exons, 1–15, their adjacent intronic sequences, and the PKD1-gene promoter region need to be isolated. We have developed a method that allows the direct sequence analysis of exons 2–15 and that is a significant step toward achievement of this aim. Comparison between PKD1-gene sequences and partial sequences from the homologous genes has allowed us to identify sequences unique to intron 1 of the PKD1 gene and to use LR-PCR to amplify a region including exons 2–15. Direct sequencing of this LR-PCR product has identified seven novel mutations in regions encoding the 16 PKD domains in 24 unrelated patients with ADPKD. The structure of these domains has recently been determined (Bycroft et al. 1999), and so the likely effect of missense mutations, in addition to the effect of nonsense and frameshift mutations, may be predicted.

Patients and Methods

Patients

DNA samples were obtained from 65 patients participating in a study to determine genetic and environmental risk factors for cerebral-aneurysm formation in ADPKD and from patients attending renal and genetics clinics. Consent for DNA storage and mutation analysis was obtained in all cases. DNA was extracted from whole blood by the Puregene kit (Gentra), according to the manufacturer's protocol. RNA was extracted from peripheral-blood lymphocytes by TRIzol Reagent (GibcoBRL). Approval for this work had been obtained from the relevant local research-ethics committees.

Sequence Alignments

Alignments of the 52-kb PKD1 genomic sequence (GenBank accession number L39891) and 127 kb of sequence containing several of the duplicated PKD1-gene sequences (GenBank accession number AC002039) was performed by DOTTER (Sonnhammer and Durbin 1995) (fig. 1).

PCR Amplification of DNA and RNA

Total RNA was reversed transcribed by EXPAND™ reverse transcriptase (Boehringer Mannheim) and random hexamers, according to the manufacture's guidelines. All LR-PCR reactions on samples from 65 patients were performed by the EXPAND™ PCR System (Boehringer Mannheim) with buffer system 3 (2.25 mM $MgCl_2$ with detergents), according to the manufacturer's protocol. Conventional PCR utilized Bio-X-Act (Bioline). A total of 200 ng of genomic DNA was used in all LR-PCR reactions. LR-PCR was performed with 50-μl reactions in 0.2-ml thin-walled PCR tubes and a PTC-200 thermocycler (MJ Research). The cycling parameters and primers used in the LR-PCR reactions are detailed in table 1. For LR-PCR reactions in buffer system 3, 500 μM of each dNTP and 300 nM of each primer were used. The primer sequences used to generate the

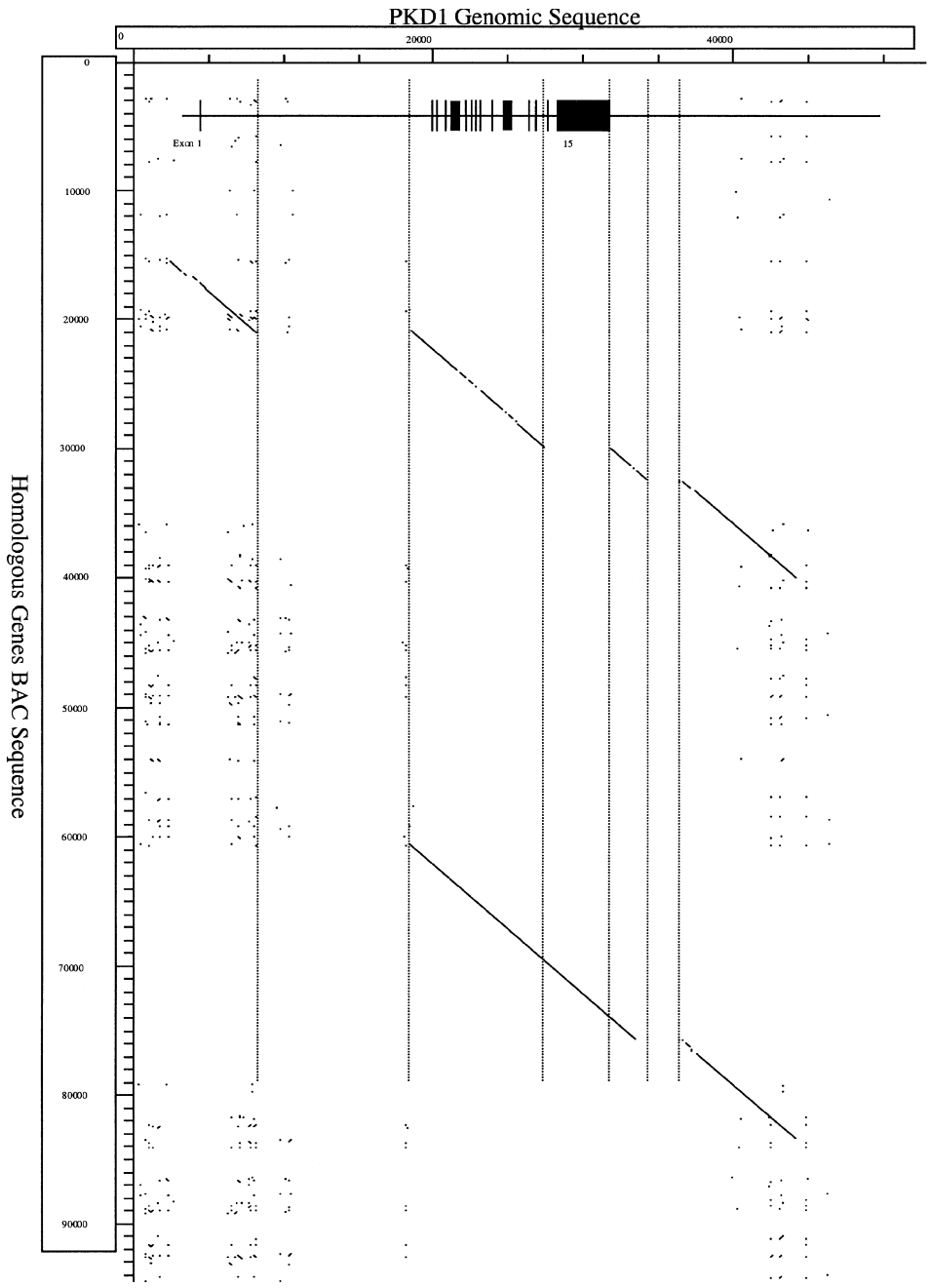

Figure 1 PKD1 genomic sequence (GenBank accession number L39891) and 127 kb of sequence from a BAC clone (GenBank accession number AC002039) containing part of the duplicated region of the PKD1 gene, aligned by DOTTER (see the Patients and Methods section). Areas of homology are defined by continuous lines. Two distinct partial copies of the PKD1 gene can be seen. Dotted vertical lines define the boundaries of regions of homology that are discontinuous with a partial genomic structure of the PKD1 gene. Two regions appear to be unique to the PKD1 gene, the largest corresponding to a substantial part of intron 1.

PCR products are listed in table 2. To generate additional sequencing templates from each patient, either LR-PCR reactions were repeated or regions of interest were amplified, by internal nested primers, from purified LR-PCR templates after 10^5-fold dilution (fig. 2).

Direct Sequencing of PCR Products

LR-PCR products from a sample of 24 patients were sequenced. LR-PCR products were purified by Qiagen™ PCR clean-up columns, and 150 ng was used in each

Table 1

LR-PCR Parameters

	5-kb Product	10-kb Product	13-kb Product
Primers	17Ext+22R	17Ext+27R	17.1L+30Ext
PCR parameters	94°C for 2 min; 10 cycles of 94°C for 10 s, 61°C for 30 s, and 68°C for 4 min; 20 cycles of 94°C for 10 s, 61°C for 30 s, and 68°C for 4 min; 68°C for 7 min; +20 s per cycle	94°C for 2 min; 10 cycles of 94°C for 10 s, 60°C for 30 s, and 68°C for 10 min; 20 cycles of 94°C for 10 s, 60°C for 30 s, and 68°C for 10 min; 68°C for 7 min, +20 s per cycle	94°C for 2 min; 10 cycles of 94°C for 10 s, 68°C for 30 s, and 68°C for 12 min; 20 cycles of 94°C for 10 s, 68°C for 30 s, and 68°C for 12 min; 68°C for 7 min, +20 s per cycle
$[Mg^{2+}]$	2.25 mM	2.25 mM	2.25 mM
Annealing temperature	61°C	60°C	68°C
Product position[a]	17049–21960	17049–27251	16947–30620
Exact size	4,911 bp	10,202 bp	13,673 bp

[a] Positions are genomic positions according to Burn et al. (1995) (GenBank accession number L39891).

sequencing reaction. PCR products were purified from agarose gels, by the QIAquick™ gel-extraction kit. All sequencing reactions were performed by the Applied Biosystems (ABI) BigDye™ Terminator Cycle Sequencing Ready Reaction Kit, according to the manufacturer's guidelines. Modifications to this protocol included total reaction volumes of 10 μl and the use of BigDye™ buffer (catalog number 361028C) to dilute the BigDye™ reaction mix two- to four-fold. Cycle sequencing parameters were 25 cycles of 95°C for 15 s, 55°C for 15 s, and 60°C for 4 min, with PCR and walking primers to span the regions of interest (table 2). Reaction products were ethanol precipitated, resuspended in formamide-loading buffer, and electrophoresed on an ABI 373A automated sequencer (fig. 3). Sequences were compared both between all individuals and with the published cDNA and genomic sequences.

Results

An alignment of the PKD1 genomic sequence against a partial sequence of the PKD1-gene homologous region revealed areas of extensive but discontinuous homology (fig. 1). Two incomplete copies of the PKD1 gene were identified in a 127-kb bacterial artificial chromosome (BAC) clone (GenBank accession number AC002039). Two regions of the PKD1 genomic sequence were not duplicated in either of the homologous copies. This suggested that the other homologous genes may have a similar structure and that these regions would be unique to the PKD1 gene. The largest region was located in intron 1, in the vicinity of nucleotide position ~8,000-18,000. Short regions close to the 3′ boundary of this area were analyzed by PCR, to determine whether they were unique to the PKD1 gene. Both the radiation hybrid, 145.19, which contains only the PKD1 gene, and the rodent-human somatic cell hybrid, 23HA, which contains only the homologous region, were used as templates, as described elsewhere (Watnick et al. 1997). Primer pair 17F and 17.5R (table 1) produced a single product of expected size (~480 bp), both with normal human genomic DNA and with DNA from cell line 145.19 only. Amplification with 17.5F and 18R produced a product from all three DNA samples, demonstrating that primer sequence 17.5F was present in the duplicated region but that primer sequence 17F was unique to the PKD1 gene. With a series of primers in the duplicated region (22R in intron 6, 27R in intron 14, and 30R at the 3′ end of exon 15), PCR products of increasing length were generated by primer 17Ext (a modified 17F primer with a higher melting temperature, for use in the LR-PCR reactions) (fig. 2). In the 13-kb LR-PCR reaction, primer 17.1L eventually replaced 17Ext, since it consistently produced a cleaner product when multiple PCR reactions were prepared. Primer-sequence modifications were also confirmed as being specific to the PKD1 gene by the same hybrid-DNA samples (fig. 4). For nested PCR reactions that were used to generate further sequencing templates, a 10^5-fold dilution of the 13-kb template was used. At this dilution, no amplification of genomic DNA carried over from the first round of PCR was detectable by primers in the single-copy region of the PKD1 gene and primers 5′ to the 13 kb template in the duplicated region (data not shown). A single 13-kb template could therefore be used to generate sufficient sequencing templates for all reactions (fig. 2).

The longest PCR product generated during this study was 13.6 kb, which extended from intron 1 to the 3′ end of exon 15 and contained ~50% of the PKD1-gene coding sequence. Exons 5 and 11-15 constitute 42% of the PKD1-gene coding sequence and encode all 16 copies of the PKD domain. The coding region and adjacent splice sites of these exons were analyzed by direct sequencing in a sample of 24 patients, to define the range of mutations in this region of the PKD1 gene and to predict the likely effect that missense mutations would have on the structure of this domain.

Table 2

Primer Sequences and PCR Conditions

PCR Product/Type and Primer (Sequence)	Genomic Position	Annealing Temperature (°C)	Mg (mM)	PCR/Sequencing	Product Length (bp)
5 kb:					
17Ext (CCC AGC GTC TCA TCT GTC TGG)[a]	17049–17069	61	2.25	LR-PCR	4,911
22R (CTG CAT CTG CAG AGC TGA CAG)	21939–21960				
10 kb:					
17Ext 27R (CAC AGG TGT AGC AGC ACT GAG)	27231–27251	60	2.25	LR-PCR	10,202
13 kb:					
17.1F (GCA TAG GCT CTC TGA GTG TCC CAC A)[a]	16947–16971	68	2.25	LR-PCR	13,673
30Ext (CCT CCA AGT AGT TGC GCT GTG ATC G)	30596–30620				
Exon 15:					
PKDD4 (GAC GCT GCC ATG GCT GTG)	26222–26239	55	2.25	PCR	4,394
30R (CAA GTA GTT GCG CTG TGA TCG)	30596–30616				
RT PCR:					
RTexon13F (ATG CCA CGC TAG CAC TGA CG)	26432–26452	58	1.5	RT-PCR	355
PKD15.2R (CGT GTT GTT GAC CTC CAG GC)	27680–27699				
ARMS PCR:					
PKDD1.F (TCC AGT GCC TCC TTT GCC)	21001–21019	52	2	ARMS-PCR	276
971WT (GCC CAG GCA GCA CAT ATC)	21260–21277				
971Mut (GCC CAG GCA GCA CAT ATA)	21260–21277				
17F (CAG CGT CTC ATC TGT CTG G)[a]	17051–17069	55	1.5	PCR	481
17.5R (GAC TAT GGC TCC GCA GGT)	17015–17532				
17.5F (AGA CCT GCG GAG CCA TAG TC)	17513–17532	61	1.5	PCR	465
18R (GCA GTG AGC TGA GAT CGC AC)	17859–17978				
PKDD1.F (TCC AGT GCC TCC TTT GCC)	21001–21019			Sequencing	
PKDD2.F (GCA CAA CCT CTC CTG CAG)	24682–24698			Sequencing	
PKDD3.1.F (CGA CAG GCT AAG GGC AGA)	25931–25948			Sequencing	
PKD15.1 (CTC AGT GCT GCT ACA CCT GT)	27231–27251			Sequencing	
PKD15.2 (GCC TGG AGG TCA ACA ACA)	27680–27699			Sequencing	
PKD15.3 (TCT TCG ACT GGA CCT TCG G)	28085–28103			Sequencing	
PKD15.4 (TGC TGC CAA TGA CTC AGC C)	28467–28485			Sequencing	
PKD15.5 (CTG TGA CCG CTG CAC GCC)	28862–28879			Sequencing	
PKD15.6 (GAC TGC ACC ATG GAC TTC)	29234–29251			Sequencing	
PKD15.7 (CAA TGT GAG CTG GTG CTG G)	29608–29626			Sequencing	
PKD15.8 (CAT CGT GGT GCT GGA GGC)	30001–30019			Sequencing	
PKDD10 (CCA CGC TTA CAA CAG CAC AG)	28653–28672			Sequencing	
PKDexon5 (GGA GCC TGT GAG TGC GGC)	20731–20748			Sequencing	
PKDexon11 (CAT GAC CGT GAG GAC GTG AT)	24312–24331			Sequencing	
PKDexon14 (CGT GAC TGC AGA GTG GAG C)	26706–26724			Sequencing	

[a] PKD1 gene–specific primer.

Mutations Identified by LR-PCR

A single 13-kb LR-PCR product (patient 8) produced two distinct fragment sizes on conventional agarose-gel electrophoresis. Further analysis of patient 8's genomic DNA by LR-PCR primers 17Ext and 22R demonstrated the presence of both a normal-size 5-kb band and an ~2-kb band (fig. 5). Direct sequencing of the smaller band revealed a deletion extending from within intron 1 to exon 5 (g18177–21076del). In a sample of 64 13-kb LR-PCR products from other patients with ADPKD, which were resolved on 0.6% agarose gels, no other altered band sizes were detected.

Mutations Identified by Direct Sequencing

Exons 5 and 11–15 were sequenced in the sample of 24 patients with ADPKD who were used in this study. The use of small-volume sequencing reactions and diluted BigDye™ terminators had no effect on sequence quality of purified PCR products, and readings >400 bp were routinely achieved (see the Patients and Methods section). All sequence variants were confirmed on both strands of two independent LR-PCR samples, to eliminate the possibility of LR-PCR artifacts. A total of 26 sequence variations were identified; 18 were silent polymorphisms, and the remaining 8 were possible phenotype-modifying mutations (tables 3 and 4).

Four of the eight possible phenotype-modifying mutations identified by direct sequencing were predicted to cause a prematurely truncated protein (table 3) and were therefore very likely to be pathological. They consisted of a 28-bp deletion (6434–6461del), a single nonsense mutation (Q1922X), a single nucleotide insertion

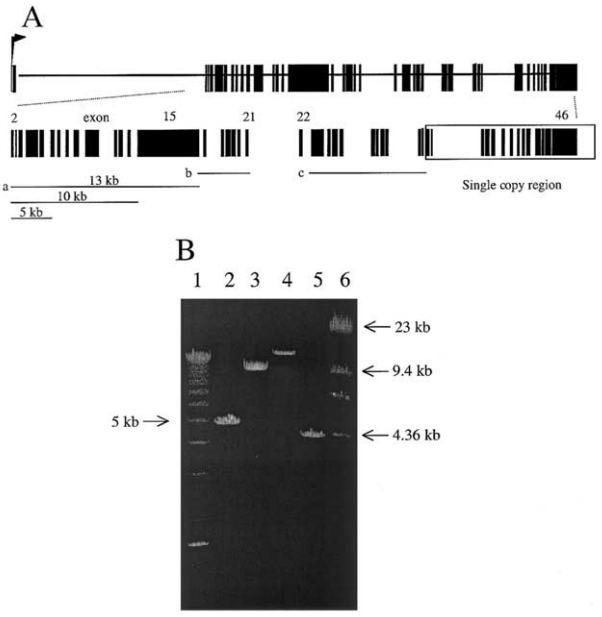

Figure 2 PKD1-gene exons 2–15, which can be specifically amplified from a unique primer in intron 1. *A*, PKD1-gene region that has been amplified by LR-PCR, shown as fragments of increasing length (line a). The shorter, 5-kb product spans exons 2–6, the 10-kb product spans exons 2–14, and the 13-kb product spans exons 2–15. Also shown are other PCR products (lines b and c) amplified from the duplicated region that have been reported by Watnick et al. (1997 [line c], 1998*b* [line b]). *B*, 0.6% Agarose gel resolving LR-PCR products. Lane 1, 1-kb marker. Lane 2, 5-kb LR-PCR product. Lane 3, 10-kb LR-PCR product. Lane 4, 13-kb LR-PCR product. Lane 5, Exon 15 amplified from a 13-kb template by primers PKDD4 and 30R. Lane 6, λ*Hin*dIII marker.

(4898–4899insT) (fig. 3), and a splice acceptor–site mutation in intron 14 (g27406G→A; IVS14−1G→A) (table 2). Frameshift mutation 4898–4899insT occurred in codon 1633 and resulted in the addition of 23 novel amino acids before a premature stop codon. Mutation 6434–6461del resulted in five novel amino acids after codon 2144, before premature termination. When both RT-PCR on total RNA isolated from peripheral-blood lymphocytes and primers in exons 13 and 15 (table 2) were used, the splicing mutation IVS14−1G→A (g27406G→A) was shown to result in intron retention (fig. 5). Direct sequence analysis of RT-PCR products demonstrated the retention of intron 14, which would result in the addition of three novel amino acids before a premature stop codon after exon 14.

The remaining four possible mutations were missense mutations—R324L, L845S, W1399R, and P1786L. On further analysis, two of these four—that is, W1399R and P1786L, located in PKD domains 8 and 12, respectively—were found to be coding polymorphisms; W1399R (4195T→C) was detected in two families with ADPKD, did not segregate with the disease phenotype in one family, and was present in 6 of 120 normal chromosomes analyzed for the presence of a newly created *Nci*I site, and P1786L (5357C→T) was confirmed in patient 41's genomic DNA by the detection of a newly created *Bsp*MI site. However, it was found in only two of three affected siblings and eventually was shown to have been inherited from the unaffected parent. It was absent from 120 normal chromosomes analyzed and was not found in any other unrelated patients with ADPKD who were studied. One of the remaining two missense mutations was located in a PKD domain (fig. 6). In patient 32, a G→T substitution was found at position 971. This resulted in a change from a basic Arg to a neutral hydrophobic Leu in PKD domain 1 (R324L). Its presence in this patient's genomic DNA was confirmed by an amplification-refractory mutation system (ARMS) PCR assay (table 2). No other family members were available for study, and the patient's three young children had not been screened for renal cysts. The final missense mutation, a T→C substitution at position 2534, predicted a change from a neutral and hydrophobic Leu to a polar Ser at position 845, six residues N-terminal to the start of PKD domain 2. It was demonstrated in the patient's genomic DNA by detection of the loss of an *Sty*I site. The mutation was not found in the patient's unaffected daughter. Both of these missense mutations were absent from 120 normal chromosomes and from all other unrelated patients with ADPKD who were studied. No patient contained more than a single potentially pathological mutation in the PKD1-gene region analyzed.

Discussion

Molecular analysis of families with ADPKD has been confined mainly to the use of informative polymorphic markers for linkage analysis. They have been used both

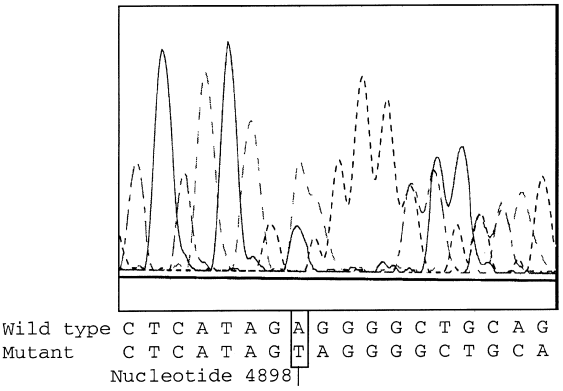

Figure 3 Direct sequence analysis of the PKD1 gene in patient 26, demonstrating a single T insertion at position 5109. The resulting frameshift introduces 23 novel amino acids after position 1633.

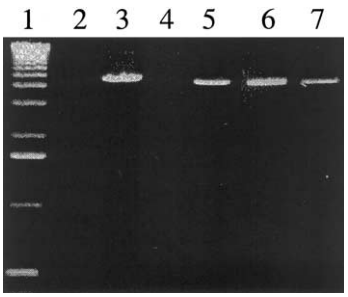

Figure 4 5-kb LR-PCR product, which is specific to the PKD1 gene. The 5-kb LR-PCR product (from use of primers 17Ext and 22R) was amplified from normal human genomic DNA as a positive control, and the specificity of the reaction was determined by use of cell lines 145.19 and 23HA (see the Results section). Lane 1, 1-kb ladder. Lane 2, No-DNA negative control. Lane 3, Human genomic DNA. Lane 4, 23HA DNA. Lane 5, 145.19 DNA. Lanes 6 and 7, 5-kb LR-PCR product generated by primers 17.5F and 22R, by use of DNA from 145.19 (lane 6) and 23HA (lane 7). This confirms that 23HA DNA can be amplified and that primer 17.5F is present in the duplicated region.

to determine whether families have linkage to either the PKD1 gene or the PKD2 gene, in predictive testing of young individuals with a normal renal ultrasound scan, and for prenatal diagnosis (Breuning et al. 1990). More-detailed genetic analysis of the PKD1 gene in affected and normal individuals has been complicated by the presence of multiple homologous copies of the PKD1 gene, which have made mutation analysis technically difficult (Harris et al. 1995). Most reported mutations are therefore located in the 3′ single-copy region of the gene. In addition, very few recurrent mutations have been described, with the majority being private (Daniells et al. 1998). This suggests that mutations are likely to be spread throughout the PKD1 gene. In this report, we have described the use of LR-PCR to identify mutations in the duplicated 5′ region of the PKD1 gene.

The specific amplification of exons 2–15 was achieved by use of a primer located in a unique region of intron 1. This was identified by alignment of the PKD1 genomic sequence with a partial sequence of the homologous region. This analysis also provided important information about the organization of the homologous genes. The BAC clone (GenBank accession number AC002039) contained two partial copies of the PKD1 gene. Numerous sequence variations were observed between these copies and the PKD1 gene, and, unexpectedly, regions of homology between the PKD1 gene and the homologous genes were clearly discontinuous (fig. 1). Several regions of the PKD1 gene were therefore potentially unique if the other homologous genes had a similar structure. These regions included a considerable part of intron 1, on the basis of which the unique primer 17Ext was designed, and the region surrounding the polypyrimidine tract in intron 21 (Burn et al. 1995). Complete sequence analysis of the rest of the homologous-gene region may permit rapid development of further PKD1 gene–specific reagents.

In previous studies, LR-PCR has been used in the analysis of the duplicated part of the PKD1 gene, to specifically amplify exons 23–34 and part of exon 15 through exon 21 (Watnick et al. 1997, 1998*b*). More than half the coding sequence, exons 1–15, has therefore not been systematically analyzed in families with ADPKD. Specific amplification of exons 2–15 now allows virtually all exons and adjacent sequences of the PKD1 gene to be examined. The specificity of the LR-PCR primers used in this study has been confirmed by use of cell lines 145.19 and 23HA, described above, which contain either the PKD1 gene alone or the homologous copies

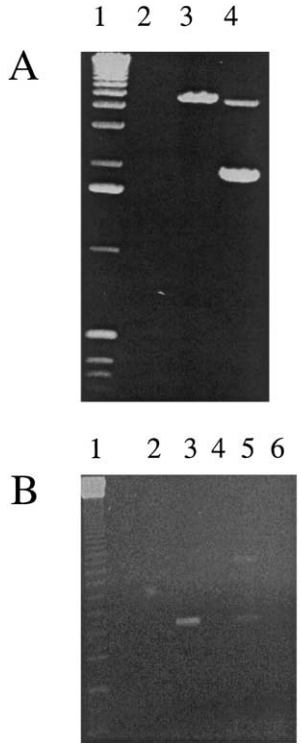

Figure 5 *A,* 5-kb LR-PCR product in patient 8, which shows a ~3-kb deletion. Amplification of genomic DNA from patient 8 (lane 4) demonstrates two bands—5 kb and 2 kb—compared with the normal, 5-kb product from the genomic control (lane 3). Lane 1, 1-kb ladder. Lane 2, No-DNA negative control. Sequencing of the 2-kb product confirmed a 2.9-kb deletion of the PKD1 gene (see the Results section). *B,* RT-PCR analysis of lymphocyte total RNA from patient 43, which shows retention of intron 14. Amplification using primers in exons 13 and 15 demonstrates a single expected 355-bp RT-PCR product in a control sample (lane 3) and two products—355 bp and 1,267 bp—in patient 43. Direct sequence analysis of the RT-PCR products confirmed that this was due to retention of intron 14 in the larger fragment. Lane 1,123-bp marker. Lane 2, No-DNA negative control. Lane 3, RT-PCR + RT of normal control. Lane 4, RT-PCR − RT in normal control. Lane 5, RT-PCR + RT in patient 43. Lane 6, RT-PCR − RT in patient 43.

Table 3

Mutations in the PKD1 Gene

Patient(s)	Mutation[a]	Location	Effect on Coding Sequence
8	g18177–21076del	Intron 1–exon 5	Frameshift after aa 72
26	4898–4899insT	Exon 15	Frameshift after aa 1633
35	6434–6461del	Exon 15	Frameshift after aa 2144
43	IVS14−1G→A	Intron 14	Frameshift after aa 1097
32	R324L	Exon 5	Arg→Leu at aa 324
33	L845S	Exon 11	Leu→Ser at aa 845
30, 38	W1399R	Exon 15	Trp→Arg at aa 1399
41	P1786L	Exon 15	Pro→Leu at aa 1786
36	Q1922X	Exon 15	Gln→stop at aa 1922

[a] All mutations except that denoted with the prefix "g" are cDNA positions, according to Hughes et al. (1995) (GenBank accession number L33234), and the genomic positions are according to Burn et al. (1995) (GenBank accession number L39891).

alone (Watnick et al. 1997). In the absence of the complete sequence of the homologous region, these reagents have been essential in the definition of many of the locus-specific reagents now used in PKD1 gene–mutation analysis. We have also analyzed PKD-domain sequences from PCR products in which amplification from the homologous region had also occurred. We consistently failed to demonstrate the same mutations that had been identified on the basis of the PKD1 gene–specific LR-PCR products, presumably because of preferential amplification of the multiple homologous genes, providing further evidence that the primers are specific to the PKD1 gene.

Attempts to amplify a region including exon 21 by the intron 1–specific primer have, so far, been unsuccessful. The size of this product, ~17 kb, is technically feasible, but certain properties of the sequence surrounding this region, such as its long polypyrimidine tract, may prevent efficient amplification. This region of intron 21 may divide the PKD1 gene into two separate halves suitable for LR-PCR amplification and mutation screening, unless an RNA-based strategy is employed.

The PCR-based approach described above has advantages over other techniques (Watnick et al. 1997). A single set of primers can be used to amplify very small quantities of genomic DNA. This produces, in one reaction, sufficient template for all future mutation screening and detection steps and allows the detection of all sequence variants. We have chosen to directly sequence the LR-PCR or nested PCR products, in preference to screening for mutations, for several reasons. Screening methods such as SSCP, which require multiple further rounds of PCR amplification of the LR-PCR templates in order to generate fragments of optimal size for analysis, are unlikely to have 100% sensitivity and cannot distinguish between silent polymorphisms, which are common in the PKD1 gene, and pathogenic mutations. The PTT will detect only those mutations that result in a truncated protein and will not detect missense mutations, which constitute an unknown proportion of PKD1-gene mutations. The PTT is also not suitable for the study of all exons, depending on either their size or the availability of RNA templates, and requires several further experimental manipulations, including the use of cell-free lysates for protein translation. Direct sequencing with BigDye™, however, can be performed directly on the LR-PCR product, uses only small amounts (~150 ng) of DNA and sequencing reagents per reaction, can be easily adapted to microtiter-plate format, and readily detects heterozygous mutations (fig. 3). The use of high-throughput automated sequencers will allow several hundred sequencing reactions to be analyzed by a single operator daily. As with other genes containing multiple exons, this will allow rapid screening of the whole coding sequence and adjacent intronic sequences, for all variants, with a high detection rate (Martin et al. 1998).

Table 4

Polymorphic Variations in the PKD1 Gene

Polymorphism[a]	Location	Effect on Coding Sequence
1023C/T	Exon 5	A341A
g25134C/T	Intron 11	
2694A/C	Exon 11	A898A
2700G/A	Exon 11	P900P
2730C/T	Exon 11	D910D
2972T/G	Exon 12	V991V
3063T/C	Exon 13	G1021G
3111A/G	Exon 13	L1037L
3372C/T	Exon 15	A1124A
3375C/T	Exon 15	S1125S
3864C/T	Exon 15	H1288H
4635G/A	Exon 15	K1545K
4665A/C	Exon 15	A1555A
4674G/A	Exon 15	T1558T
5172C/T	Exon 15	A1724A
5359C/T	Exon 15	L1787L
5485C/T	Exon 15	L1829L
5763G/A	Exon 15	L1921L

[a] Positions are as defined in the footnote to table 3.

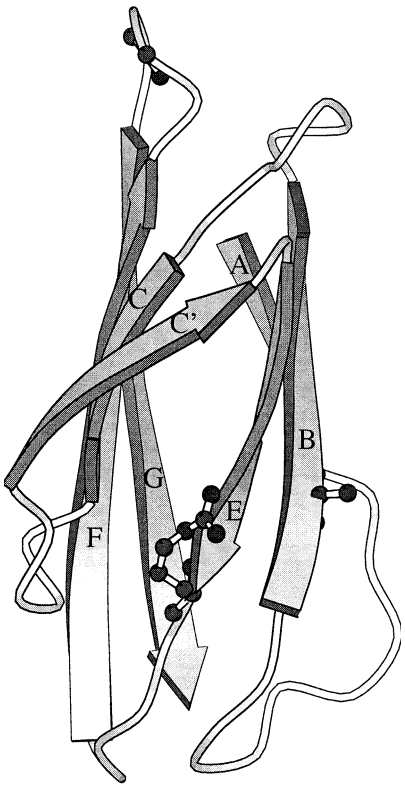

Figure 6 PKD domain, which is a β sandwich composed of two sheets, one containing A, B, and E strands and the other containing G, F, C, and C′ strands. The positions of R324L (domain 1, E strand), P1786L (domain 12, FG turn), and W1399R (domain 8, B strand) are shown superimposed on the structure of PKD domain 1. All amino acid changes can be seen to affect surface residues. This figure was prepared by Molscript (Kraulis 1991).

Because of the ability to use small (10 μl) sequencing-reaction volumes and diluted sequencing reagents, the costs and time taken to detect mutations, are considerably less for this method than they are those for methods such as PTT. During this study, reagent costs for LR-PCR and DNA purification that produced sufficient template for six to eight sequencing reactions were 2£ ($3.20)/patient sample. Further sequencing reactions were performed on template generated by nested primer sequences and LR-PCR template DNA diluted 10^5-fold (table 2). Therefore, a single LR-PCR reaction generated sufficient DNA for all future applications. Costs for each sequencing reaction were calculated to be ~1£ ($1.60), excluding sequencer-related costs. However, the main advantage of this technique is the detection of all sequence variants in the region being analyzed. This study has demonstrated a mutation-detection rate of 29% in a sample of 24 patients. If 85% of these patients have a mutation in the PKD1 gene, then the detection rate was 7/20, or 35%, when 42% of the coding region was analyzed.

The generation of a PKD1 gene–specific LR-PCR product spanning exons 2–15 now allows a large part of the duplicated region of the PKD gene to be screened for mutations. Five of the mutations identified were either frameshift or stop mutations and therefore were clearly pathogenic. In the case of the splicing mutation, IVS14−1G→A, the G→A substitution abolishes the 3′ splice-acceptor site of intron 14 and would be predicted to lead to aberrant splicing. The possible effects of this mutation include the use of alternative cryptic splice sites in either intron 14 or exon 15 (several of which can be predicted), intron retention, or exon skipping. RT-PCR analysis of peripheral blood–lymphocyte RNA has shown that retention of intron 14 occurs, producing a premature stop codon after three novel amino acids are added to exon 14. It has not been possible to determine whether the same or additional kidney-specific splicing events occur.

Although the presence of a truncating mutation is clearly pathogenic, a determination of the significance of a missense mutation is more difficult. Established criteria include the absence of other significant mutations elsewhere in the gene, segregation of the mutation with the disease phenotype, alteration of conserved amino acid residues to those with different physical properties, occurrence on <1% of normal alleles, and the demonstration of altered function (Cotton and Scriver 1998). Clearly, not all of these criteria can be fulfilled for these missense mutations, and hence their designation as pathological may remain uncertain until the rest of the gene can be completely screened and until definitive functional assays can be performed. Since the function of the extracellular region of polycystin—and, in particular, that of the PKD domains—is unknown, the effect of the missense mutations described in this report can be deduced only from both their location in the PKD domain and their likely effect on structure or ligand binding.

The structure of the PKD domain has recently been determined (Bycroft et al. 1999). It is a novel protein domain consisting of a β-sandwich or immunoglobulin (Ig) fold (fig. 6). Two β-sheets, one of which has three strands (A, B, and E) and the other of which has one of four strands (G, F, C, and C′), are packed face-to-face with a well-defined hydrophobic core centered around a conserved tryptophan. This fold is common to many proteins of diverse function. The PKD domain represents a novel protein domain that may be involved in protein-protein interactions. Comparison with the binding features of other Ig-fold proteins, including members of the Ig superfamily, reveals that each part of the surface of the domain, loops and sheets, can be used for interaction with other proteins (Bork et al. 1994). Therefore, each face of the PKD domain may be involved in either ligand binding or interactions with adjacent PKD domains,

since 15 of the 16 domains in polycystin are arranged in tandem.

Mutations in these domains may be divided into those that act mainly by changing conformation or denaturing their domain and those that alter the domain's surface properties, as has been suggested with regard to other Ig-fold proteins (Bateman et al. 1996). Both the coding polymorphisms and the missense mutation identified in the PKD domains occur on surface residues. They are therefore unlikely to have a major effect on the structure of each domain. The W1399R mutation, despite the change from an aromatic hydrophobic residue to a basic one in the central part of the β strand of PKD domain 8, does not segregate with ADPKD and is found at a low frequency in the normal population. This position in other PKD domains in man and *Fugu* is not conserved and may be occupied by an Arg residue or other basic, polar, or hydrophobic residues (Sandford et al. 1997); it is therefore unlikely to have major functional importance. It is possible that this change may have a modifying effect on disease expression, if, for example, it alters ligand-binding affinities. The Pro→Leu change at position 1786 in PKD domain 12 also does not segregate with ADPKD in the family in which it was detected and is therefore not disease causing. It occurs in the turn between the F and G strands, which, in all human PKD domains, comprises only two residues. This position is not conserved during evolution, and Leu occupies this position in other PKD domains. It is a rare polymorphism, which, again, may exert some effect on disease expression, by altering potential PKD-protein interactions.

Mutation R324L in PKD domain 1 is located on the E strand. This domain is unique, since it occurs in isolation and is therefore unlikely to participate in interactions with other PKD domains. It is the only domain that has a basic residue in this position. The same position in the *Fugu* PKD domain 1 is also occupied by a basic residue, Lys, further suggesting the functional importance of this position. The change from a basic residue to a cyclic, neutral hydrophobic residue is more likely to interfere with ligand binding than with the conformation of the domain itself. The E strand is therefore a potential site of protein-protein interactions on the PKD domain.

Mutation L845S occurs six residues N-terminal to the start of PKD domain 2. It is therefore not possible to determine whether it has any structural or binding significance, but the change from a hydrophobic to a polar residue in a position that is also conserved in *Fugu* polycystin suggests that it is important for protein structure or function.

Although these missense mutations fulfill most of the criteria required in order for them to be designated as pathological, final confirmation of their role in the pathogenesis of ADPKD will depend on the elucidation of the function of both polycystin-1 and the PKD domains and on investigation of their effects in a functional assay. Similarly, the possibility that some of the silent polymorphisms detected, which generate potential cryptic splice sites, might be pathological will await the screening, for other possible pathogenic mutations, of the rest of the gene in these individuals.

Gene conversion has been shown to be a cause of mutations in the PKD1 gene (Watnick et al. 1998*a*). Of the four missense changes found in this study, only mutation 2534T→C was found to present in the homologous genes. This was confirmed by BAC sequence AC002039, for comparison, and by DNA from cell line 23HA.

The PKD1 gene still represents a considerable challenge for the detection of pathogenic mutations. This challenge is mainly due to its large size, since the presence of multiple copies no longer presents an obstacle to the analysis of the vast majority of its coding sequence.

Acknowledgments

We are grateful to all the patients, for their contribution to this study. We would also like to thank Sue Tebbs and Kathy Barnes, for the collection of clinical information and DNA samples; Dr. Peter Harris, for the kind gift of DNA from cell line 145.19; and Drs. Mark Bycroft and Jane Clarke, for helpful discussions on the PKD structure. This work was funded by grants from the National Kidney Research Fund, The Stroke Association, Addenbrooke's Hospital NHS Trust, The Addenbrooke's Hospital Kidney Patients Association, The Medical Research Council, and The Wellcome Trust. R.S. is a Wellcome Trust Senior Fellow in Clinical Research and a former MRC Clinician Scientist.

Electronic-Database Information

GenBank, http://www.ncbi.nlm.nih.gov/Web/Genbank (for BAC clone [AC002039] and for PKD1-gene cDNA nucleotide positions [L33234] and genomic nucleotide positions [L39891])

Human Gene Mutation Database Cardiff, The, http://www.uwcm.ac.uk/uwcm/mg/hgmd0.html

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for ADPKD [MIM 173900])

References

Bateman A, Jouet M, Macfarlane J, Du JS, Kenwrick S, Chothia C (1996) Outline structure of the human L1 cell-adhesion molecule and the sites where mutations cause neurological disorders. EMBO J 15:6050–6059

Bork P, Holm L Sander C (1994) The immunoglobulin fold: structural classification, sequence patterns and common core. J Mol Biol 242:309–320

Breuning MH, Snijdewint FG, Dauwerse JG, Saris JJ, Bakker E, Pearson PL, van Ommen GJ (1990) Two step procedure

for early diagnosis of polycystic kidney disease with polymorphic DNA markers on both sides of the gene. J Med Genet 27:614–617

Burn TC, Connors TD, Dackowski WR, Petry LR, Van Raay TJ, Millholland JM, Venet M, et al (1995) Analysis of the genomic sequence for the autosomal dominant polycystic kidney-disease (PKD1) gene predicts the presence of a leucine-rich repeat: The American PKD1 Consortium. Hum Mol Genet 4:575–582

Bycroft M, Bateman A, Clarke J, Hamill SJ, Sandford R, Thomas RL, Chothia C (1999) The structure of a PKD domain from polycystin-1: implications for polycystic kidney disease. EMBO J 18:297–305

European Polycystic Kidney Disease Consortium, The (1994) The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16: The European Polycystic Kidney Disease Consortium. Cell 77:881–894

Cotton RGH, Scriver CR (1998) Proof of disease causing mutation. Hum Mutat 12:1–3

Daniells C, Maheshwar M, Lazarou L, Davies F, Coles G, Ravine D (1998) Novel and recurrent mutations in the PKD1 (polycystic kidney disease) gene. Hum Genet 102:216–220

Harris PC, Ward CJ, Peral B, Hughes J (1995) Autosomal-dominant polycystic kidney disease: molecular analysis. Hum Mol Genet 4:1745–1749

Hateboer N, Dijk M, Bogdanova N, Coto E, Saggar-Malik AK, Millan JLS, Torra R, et al (1999) Comparison of phenotypes of polycystic kidney disease types 1 and 2. Lancet 353:103–107

Hughes J, Ward CJ, Peral B, Aspinwall R, Clark K, Sanmillan JL, Gamble V, et al (1995) The polycystic kidney-disease-1 (Pkd1) gene encodes a novel protein with multiple cell recognition domains. Nat Genet 10:151–160

Kraulis PJ (1991) Molescript—a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 24:946–950

Martin P, Heiskari N, Zhou J, Leinonen A, Tumelius T, Hertz JM, Barker D, et al (1998) High mutation detection rate in the COL4A5 collagen gene in suspected Alport syndrome using PCR and direct DNA sequencing. J Am Soc Nephrol 9:2291–2301

Mochizuki T, Wu G, Hayashi T, Xenophontos SL, Veldhuisen B, Saris JJ, Reynolds DM, et al (1996) PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. Science 272:1339–1342

Paterson AD, Pei Y (1998) Is there a third gene for autosomal dominant polycystic kidney disease? Kidney Int 54: 1759–1761

Pei Y, He N, Wang K, Kasenda M, Paterson AD, Chan G, Liang Y, et al (1998) A spectrum of mutations in the polycystic kidney disease-2 (PKD2) gene from eight Canadian kindreds. J Am Soc Nephrol 9:1853–1860

Peral B, Gamble V, Strong C, Ong ACM, Sloane-Stanley J, Zerres K, Winearls CG, et al (1997) Identification of mutations in the duplicated region of the polycystic kidney disease 1 gene (*PKD1*) by a novel approach. Am J Hum Genet 60:1399–1410

Peral B, Ong AC, San Millan JL, Gamble V, Rees L, Harris PC (1996*a*) A stable, nonsense mutation associated with a case of infantile onset polycystic kidney disease 1 (PKD1). Hum Mol Genet 5:539–542

Peral B, San Millán JL, Ong ACM, Gamble V, Ward CJ, Strong C, Harris PC (1996*b*) Screening the 3′ region of the polycystic kidney disease 1 (*PKD1*) gene reveals six novel mutations. Am J Hum Genet 58:86–96

Peters DJM, Sandkuijl LA (1992) Genetic heterogeneity of polycystic kidney disease in Europe. In: Breuning MH, Devoto M, Romeo G (eds) Contributions to nephrology. Vol 97: Polycystic kidney disease. Karger, Basel, pp 128–139

Qian F, Watnick TJ, Onuchic LF, Germino GG (1996) The molecular basis of focal cyst formation in human autosomal dominant polycystic kidney disease type I. Cell 87:979–987

Roelfsema JH, Peters DJ, Breuning MH (1996) Detection of translation terminating mutations in the PKD1 gene. Nephrol Dial Transplant 11:5–9

Roelfsema JH, Spruit L, Saris JJ, Chang P, Pirson Y, van Ommen G-JB, Peters DJM, et al (1997) Mutation detection in the repeated part of the PKD1 gene. Am J Hum Genet 61: 1044–1052

Sandford R, Sgotto B, Aparicio S, Brenner S, Vaudin M, Wilson RK, Chissoe S, et al (1997) Comparative analysis of the polycystic kidney disease 1 (PKD1) gene reveals an integral membrane glycoprotein with multiple evolutionary conserved domains. Hum Mol Genet 6:1483–1489

Sonnhammer EL, Durbin R (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167:1–10

Veldhuisen B, Saris JJ, de Haij S, Hayashi T, Reynolds DM, Mochizuki T, Elles R, et al (1997) A spectrum of mutations in the second gene for autosomal dominant polycystic kidney disease (PKD2). Am J Hum Genet 61:547–555

Watnick TJ, Gandolph MA, Weber H, Neumann HP, Germino GG (1998*a*) Gene conversion is a likely cause of mutation in PKD1. Hum Mol Genet 7:1239–1243

Watnick TJ, Piontek KB, Cordal TM, Weber H, Gandolph MA, Qian F, Lens XM, et al (1997) An unusual pattern of mutation in the duplicated portion of PKD1 is revealed by use of a novel strategy for mutation detection. Hum Mol Genet 6:1473–1481

Watnick TJ, Torres VE, Gandolph MA, Qian F, Onuchic LF, Klinger KW, Landes G, et al (1998*b*) Somatic mutation in individual liver cysts supports a two-hit model of cystogenesis in autosomal dominant polycystic kidney disease. Mol Cell 2:247–251

Watson ML, Torres VE (ed) (1996) Polycystic kidney disease. In: Oxford clinical nephrology series. Oxford University Press, Oxford