Am. J. Hum. Genet. 65:1561–1571, 1999

Mutation Detection of *PKD1* Identifies a Novel Mutation Common to Three Families with Aneurysms and/or Very-Early-Onset Disease

Terry Watnick,[1] Bunyong Phakdeekitcharoen,[1] Ann Johnson,[3] Michael Gandolph,[1] Mei Wang,[4] Gary Briefel,[2] Katherine W. Klinger,[5] William Kimberling,[4] Patricia Gabow,[3] and Gregory G. Germino[1]

[1]Johns Hopkins University School of Medicine, Division of Nephrology, and [2]Johns Hopkins-Bayview Hospital, Division of Nephrology, Baltimore; [3]University of Colorado Health Sciences Center, Polycystic Kidney Disease Research Group, Denver; [4]Department of Genetics, Center for Hereditary and Communication Disorders, Boys Town National Research Hospital, Omaha; [5]Genzyme Corporation, Framingham, MA

Summary

It is known that several of the most severe complications of autosomal-dominant polycystic kidney disease, such as intracranial aneurysms, cluster in families. There have been no studies reported to date, however, that have attempted to correlate severely affected pedigrees with a particular genotype. Until recently, in fact, mutation detection for most of the *PKD1* gene was virtually impossible because of the presence of several highly homologous loci also located on chromosome 16. In this report we describe a cluster of 4 bp in exon 15 that are unique to *PKD1*. Forward and reverse *PKD1*-specific primers were designed in this location to amplify regions of the gene from exons 11–21 by use of long-range PCR. The two templates described were used to analyze 35 pedigrees selected for study because they included individuals with either intracranial aneurysms and/or very-early-onset disease. We identified eight novel truncating mutations, two missense mutations not found in a panel of controls, and several informative polymorphisms. Many of the polymorphisms were also present in the homologous loci, supporting the idea that they may serve as a reservoir for genetic variability in the *PKD1* gene. Surprisingly, we found that three independently ascertained pedigrees had an identical 2-bp deletion in exon 15. This raises the possibility that particular genotypes may be associated with more-severe disease.

Received April 22, 1999; accepted October 1, 1999; electronically published November 12, 1999.

Address for correspondence and reprints: Dr. Gregory G. Germino, Division of Nephrology, Johns Hopkins University School of Medicine, Ross 970, 720 Rutland Avenue, Baltimore, MD 21205. E-mail: ggermino@welch.jhu.edu

0002-9297/1999/6506-0011$02.00

Introduction

PKD1, the gene mutated in the most common form of autosomal-dominant polycystic kidney disease (ADPKD [MIM 173900]) is predicted to encode a novel 460-kD membrane protein of unknown function (American PKD1 Consortium 1995; Hughes et al. 1995; International Polycystic Kidney Disease Consortium 1995). Its gene product has a number of domains found in other proteins that function as receptors for either extracellular-matrix or other cell-surface proteins, prompting speculation that PKD1 plays a similar role in renal and biliary epithelial cells. No functional studies have yet been reported that confirm these predictions, however.

One strategy commonly used to gain insights into the properties of a disease-associated gene is to define the pattern of mutations present in affected individuals. This approach can identify functionally important domains that can be targeted for experimental investigation. Another advantage of this strategy is that correlation of genotypes with detailed clinical evaluations may provide clinically useful information. In the case of ADPKD, it has been previously shown that the most severe complications of the disease, such as intracranial aneurysms, aortic dissections, and early-onset disease, often cluster in families (Chapman et al. 1992; Fick et al. 1993; Zerres et al. 1993). It is not known, however, whether the familial clustering is the result of particular types of mutations that are associated with specific phenotypic features or of the effects of other genetic factors encoded by distinct loci. None of the published *PKD1* mutations, except for the small subset that also disrupts the adjacent tuberous sclerosis 2 gene (*TSC2*), appear to be associated with a particular phenotype (Brook-Carter et al. 1994). It should be noted, however, that none of the studies have been performed on a large panel of patients who suffer from the most severe forms of disease and that relatively few mutations have been described (summarized in the Human Gene Mutation Database).

A number of features of the *PKD1* gene's structure

have greatly hindered efforts to complete genotype/phenotype analyses. *PKD1* is encoded by 46 exons that yield a coding sequence >13 kb (Hughes et al. 1995 [GenBank accession numbers L33234 and L39891 for *PKD1* cDNA and genomic sequence, respectively]). Approximately 70% of the gene's length (exons 1–34; fig. 1) is replicated elsewhere on chromosome 16, in multiple highly homologous copies that are also transcribed (Germino et al. 1992; European Polycystic Kidney Disease Consortium 1994). The sequence identity in the replicated segments is >95% and includes both exonic and intronic sequences. Thus, mutation detection in the replicated region of the gene depends on the identification of locus-specific primers that can selectively amplify *PKD1* but not its homologues.

Both our group and others have previously used a unique sequence in the 3′ portion of *PKD1* to design anchor primers for long-range amplification of exons 23–34 from both cDNA and genomic DNA (Peral et al. 1997; Roelfsema et al. 1997; Watnick et al. 1997). These gene-specific templates have been successfully employed for mutation analysis by use of a number of techniques, including heteroduplex analysis and single-strand conformation analysis (SSCA). Amplification of longer templates from genomic DNA has been difficult because of the presence of a long, 2.5-kb polypyrimidine tract in intron 21 and a shorter element in intron 22 (Van Raay et al. 1996). The high degree of sequence identity between *PKD1* and its homologues has made it difficult to design additional *PKD1*-specific primers that could be used to anchor amplification of the 5′ end of the gene (Watnick et al. 1998*a*). Another complicating factor is the additional observation that not all of the homologues (at least three) differ from *PKD1* in exactly the same location.

In this study we report a cluster of 4 bp in exon 15 that are unique to *PKD1*. Both forward and reverse primers were designed to exploit these differences and were used to amplify *PKD1*-specific products of ~3.5 (5′ middle range [MR]) and ~6.0 kb (5′ longer range [LR]). Together, these templates contain ~5.5 kb of coding sequence. A series of control reactions was employed to confirm the specificity of 5′ MR and 5′ LR and to demonstrate that they can serve as templates for nested amplification of gene-specific fragments if appropriately diluted. We used this approach to evaluate genomic DNA samples from 35 pedigrees with individuals having ADPKD along with either early-onset disease and/or intracranial aneurysms. We identified eight pathogenic sequence changes, several variants associated with disease and not found in normal individuals, and a variety of useful polymorphisms. Three independently ascertained and unrelated pedigrees were found to share an identical pathogenic mutation.

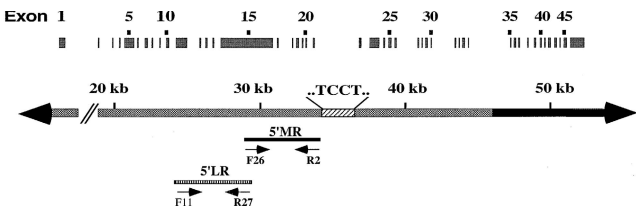

Figure 1 *PKD1* gene, which contains 46 exons and is bisected by a polypyrimidine tract of ~2.5 kb (*diagonally striped box*). The replicated portion of the gene begins with exon 1 and is thought to end in exon 34 (*stippled box* [European Polycystic Kidney Disease Consortium 1994]). Two locus-specific PCR products, 5′ MR (*black*) and 5′ LR (*vertically striped*) were used for mutation detection. The approximate position of the two templates, along with the primers used to amplify them, is indicated.

Subjects and Methods

Patient Recruitment

The University of Colorado Polycystic Kidney Research Study has 806 affected individuals who have participated in the ongoing natural-history study of ADPKD. All patients who participated in the study underwent a formalized history and physical examination and complete abdominal ultrasonography, as described elsewhere (Gabow et al. 1989). A subset of patients also participated in the study of intracranial aneurysms in ADPKD, as described elsewhere (Chapman et al. 1992). An additional set of patients was recruited from dialysis centers and nephrology clinics in Baltimore. All patients who participated in the studies had blood obtained for genetic analysis. Informed consent was obtained from all patients, in accordance with institutional guidelines at University of Colorado and Johns Hopkins University. For this study, DNA samples were utilized from patients who were diagnosed during the 1st year of life (very early onset) and from individuals who had a documented intracranial aneurysm, whether ruptured or unruptured (Chapman et al. 1992, Fick et al. 1993).

Long-Range PCR

DNA was isolated from whole blood by use of a PE Biosystems 340A nucleic-acid–extraction machine or the PureGene DNA-extraction kit (Gentra). Three hundred nanograms of genomic DNA was then used as template for amplification of a ~3.5-kb product (5′ MR), by use of primers F26 (5′-AGCGCAACTACTTGGAGGCCC-3′, genomic position 30606–30626) and R2LR (5′-GCAGGGTGAGCAGGTGGGGCCATCCTAC-3′, genomic position 33957–33984). 5′ LR was amplified from 400 ng of genomic template by use of primers F11 (5′-TGCCCCTGGGAGACCAACGATAC-3′, genomic position 24911–24933) and R27 (5′-GTCAACGTGGG-

CCTCCAAGT-3′, genomic position 30615–30634). For the cell lines 145.19 and N23HA, 400 ng of genomic DNA, prepared by use of the Puregene kit, was always used as template. PCR was performed as follows for 5′ MR: denaturation at 95°C for 3 min, 35 cycles at 95°C for 20 s and 70°C for 4 min and 30 s, and a final extension at 72°C for 10 min. The total PCR volume was 50 μl, with 4 units of r*Tth* DNA polymerase (XL; PE Biosystems) and a final MgOAC2 concentration of 1.1 mM. For 5′ LR, the following PCR conditions were employed: denaturation at 95°C for 3 min and 15 s, 35 cycles at 95°C for 20 s and 68°C for 7 min, and a final extension at 72°C for 10 min. The PCR-reaction mix was the same as that described for 5′ MR, except that the final MgOAC concentration was 1.0 mM. A hot-start protocol as recommended by the manufacturer was used for the first cycle of amplification. Products of the long-range PCR reaction were run on a 1% agarose gel to confirm that the reaction was successful before the next step was undertaken. The specificity of long-range products was evaluated by testing for the presence of a PCR product containing exon 32. The primers and conditions for this PCR reaction have been published elsewhere (Watnick et al. 1997).

SSCA Analysis

The long-range products 5′ MR and 5′ LR were diluted serially to $1 : 10^{-4}$, and 2 μl of diluted template was used as template for all subsequent PCR reactions. Intron-based primers were designed for each exon with the exception of exon 15, which required amplification in 18 separate overlapping fragments. Only a portion of exon 11 is contained in 5′ LR. In total, 25 different primer pairs were designed. The sequences of these primers and the PCR-reaction conditions are summarized in table 1. The total PCR volume was 20 μl, with 2 units of *Taq* DNA polymerase (Boehringer Mannheim), 0.2 μl dCTP, and a final MgCl2 concentration of 1.5 mM.

SSCA analysis was performed by use of 8% polyacrylamide gels with 5% glycerol added. The radiolabeled PCR products were diluted with loading buffer, were denatured by heating at 95°C for 5 min, and then were placed on ice prior to being loaded and run on the gel at room temperature. Gels were run at 400 V overnight, dried, and placed on X-Omat XAR film (Kodak) at room temperature. PCR products with variants were reamplified from the diluted template and were purified from an agarose gel slice by means of Qiajex and the manufacturer's protocol (Qiagen) or, alternatively, directly with Centricon-100 columns (Amicon). When possible, mutant SSCP bands were also excised from the agarose gel, and the DNA was eluted and used for PCR amplification and then for sequencing. In the case of certain sequence variants containing insertions or deletions, the PCR product was cloned into pCRII (TA Cloning Kit; Invitrogen) prior to being sequenced. PCR products and clones were sequenced with ^{32}P cycle sequencing by use of the Thermosequenase kit (Amersham). When possible, sequence differences were confirmed by restriction digestion of PCR products by use of restriction enzymes whose recognition sites were altered by the sequence change in question. In several instances, the change did not alter a restriction site, so mismatch oligonucleotides were designed that introduce a new restriction site when combined with the point mutation in question (Shoffner 1998). The primers were designed to create a restriction site when the alteration in question was present. For the G5689A mutation in UC109 (table 2), the primers 15-6StopF (5′-CGAGCCATTTACCACCCATAGCTTCC-3′) and 15-6StopR (5′-ATTGGTGCCCGTGGCCAGCAGCGC-3′) were used with the enzyme *Eco*47III. For the polymorphism in exon 18, the primer 18PolyF (5′-GCTCTTGCCGCCTCTTCCCA-3′) was used in combination with 18R (table 1) and the enzyme *Nsi*I. Pathogenic mutations were confirmed in family members, when possible, by SSCA or by restriction digest.

Results

Locus-Specific Templates

In the process of assembly of a cDNA contig, a series of clones was identified that, in 4 bp clustered in a short stretch of exon 15 (fig. 2), differed from *PKD1*. Because more than one clone contained these sequence differences, we hypothesized that these clones were likely to be derived from the homologous loci and not from *PKD1*. A subsequent search of the GenBank database identified the sequence of a bacterial artificial chromosome (BAC) that maps to 16p13.1 and contains two homologous copies of *PKD1* (GenBank accession number AC002039). Although one copy lacked most of exon 15, the other copy contained the same 4-bp substitutions as did the cDNA clones presumed to have originated from the homologous loci. To take advantage of this *PKD1*-specific sequence, forward (F26) and reverse (R27) primers were designed to incorporate all 4-bp mismatches with one at the 3′ terminus, to maximize specificity (fig. 2). Both primers work well for long-range PCR and can be used to amplify products of ~3.5 (5′ MR) and ~6.0 kb (5′ LR) from genomic DNA (Watnick et al. 1998*b*). These templates span part of exons 11–21 and include ~5.5 kb of coding sequence (fig. 1 and Subjects and Methods). The two templates overlap at the site of the *PKD1*-specific primer.

To prove that the primers F26 and R27 are *PKD1* specific, we used two cell lines that have been described elsewhere. In brief, N23HA is a rodent-human somatic-

Table 1

Primers Used for Mutation Detection

Exon	Primers	Fragment Size (bp)	Annealing Temperature (°C)
5′ LR:			
11	F11: 5′-TGCCCCTGGGAGACCAACGATAC-3′	303	67
	11R1: 5′-GGCTGCTGCCCTCACTGGGAAG-3′		
12	12F: 5′-GAGGCGACAGGCTAAGGG-3′	286	64
	12R: 5′-CATGAAGCAGAGCAGAAGGC-3′		
13	13F: 5′-TGGAGGGAGGGACGCCAATC-3′	308	67
	13R: 5′-GAGGCTGGGGCTGGGACAAG-3′		
14	14F: 5′-CCCGGTTCACTCACTGCG-3′	220	64
	14R: 5′-CCGTGCTCAGAGCCTGAAAG-3′		
15	15F16: 5′-CGGGTGGGGAGCAGGTGG-3′	280	67
	15R16: 5′-GCTCTGGGTCAGGACAGGGGA-3′		
	15F15: 5′-CGCCTGGGGGTGTTCTTT-3′	270	64
	15R15: 5′-ACGTGATGTTGTCGCCCG-3′		
	15F14: 5′-GCCCCCGTGGTGGTCAGC-3′	250	67
	15R14: 5′-CAGGCTGCGTGGGGATGC-3′		
	15F13: 5′-CTGGAGGTGCTGCGCGTT-3′	256	67
	15R13: 5′-CTGGCTCCACGCAGATGC-3′		
	15F12: 5′-CGTGAACAGGGCGCATTA-3′	270	65
	15R12: 5′-GCAGCAGAGATGTTGTTGGAC-3′		
	15F11: 5′-CCAGGCTCCTATCTTGTGACA-3′	259	60
	15R11: 5′-TGAAGTCACCTGTGCTGTTGT-3′		
	15F10: 5’-CTACCTGTGGGATCTGGGG-3’	217	67
	15R10: 5′-TGCTGAAGCTCACGCTCC-3′		
	15F9: 5′-GGGCTCGTCGTCAATGCAAG-3′	267	67
	15R9: 5′-CACCACCTGCAGCCCCTCTA-3′		
	15F8: 5′-CCGCCCAGGACAGCATCTTC-3′	261	64
	15R8: 5′-CGCTGCCCAGCATGTTGG-3′		
	15F7: 5′-CGGCAAAGGCTTCTCGCTC-3′	288	64
	15R7: 5′-CCGGGTGTGGGGAAGCTATG-3′		
	15F6: 5′-CGAGCCATTTACCACCCATAG-3′	231	65
	15R6: 5′-GCCCAGCACCAGCTCACAT-3′		
	15F5: 5′-CCACGGGCACCAATGTGAG-3′	251	64
	15R5: 5′-GGCAGCCAGCAGGATCTGAA-3′		
	15F4: 5′-CAGCAGCAAGGTGGTGGC-3′	333	67
	15R4: 5′-GCGTAGGCGACCCGAGAG-3′		
	15F3: 5′-ACGGGCACTGAGAGGAACTTC-3′	206	64
	15R3: 5′-ACCAGCGTGCGGTTCTCACT-3′		
	15F2: 5′-GCCGCGACGTCACCTACAC-3′	265	67
	15R2: 5′-TCGGCCCTGGGCTCATCT-3′		
	15F1: 5′-GTCGCCAGGGCAGGACACAG-3′	228	68
	R27: 5′-AGGTCAACGTGGGCCTCCAA-3′		
5′ MR:			
15	15F1-1: 5′-ACTTGGAGGCCCACGTTGACC-3′	276	69
	15R1-1: 5′-TGATGGGCACCAGGCGCTC-3′		
	15F1-2: 5′-CATCCAGGCCAATGTGACGGT-3′	266	64
	15R1-2: 5′-CCTGGTGGCAAGCTGGGTGTT-3′		
16	16F: 5′-TAAAACTGGATGGGGCTCTC-3′	294	56
	16R: 5′-GGCCTCCACCAGCACTAA-3′		
17	17F: 5′-GGGTCCCCCAGTCCTTCCAG-3′	244	67
	17R: 5′-TCCCCAGCCCGCCCACA-3′		
18	18F: 5′-GCCCCCTCACCACCCCTTCT-3′	342	67
	18R: 5′-TCCCGCTGCTCCCCCCAC-3′		
19	19F: 5′-GATGCCGTGGGGACCGTC-3′	285	67
	19R: 5′-GTGAGCAGGTGGCAGTCTCG-3′		
20	20F: 5′-CCACCCCCTCTGCTCGTAGGT-3′	232	64
	20R: 5′-GGTCCCAAGCACGCATGCA-3′		
21	21F: 5′-TGCCGGCCTCCTGCGCTGCTGA-3′	232	67
	TWR2: 5′-GTAGGATGGCCCCACCTGCT-CACCCTGC-3′		

cell hybrid that contains only the homologous loci, whereas 145.19 is a radiation hybrid that contains only *PKD1* (Germino et al. 1990; Ceccherini et al. 1992; European Polycystic Kidney Disease consortium 1994). As shown in figure 3, 5′ MR and 5′ LR can be amplified from the DNA of 145.19 but not from that of N23HA. The DNA of N23HA does, however, support the amplification of similar-sized PCR products when appropriate primers are used (data not shown). 5′ MR and 5′ LR can be used for mutation analysis of exons 11–21 by use of techniques similar to those described for exons 23–34 (Watnick et al. 1997); that is, the templates can be used for nested amplification once they have been adequately diluted. A dilution $\geqslant 10^4$ is required to remove contamination of total genomic DNA used in the initial amplification reaction (fig. 4).

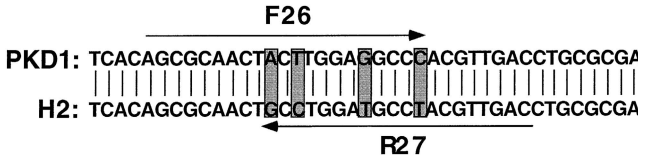

Figure 2 Cluster of 4 bp unique to *PKD1*. An alignment of exon 15 sequence between *PKD1* (accession number L33243) and one of the homologous loci is shown. The second homologue present in a BAC that maps to 16p13.11 (accession number AC002039) lacks most of exon 15, including this region. The 4 bp that differ between the two sequences are boxed and shaded. Forward (F26) and reverse (R27) *PKD1* primers were designed to include all four mismatches, including one at the 3′ end of each primer. The location of each primer is indicated by an arrow.

Probable Pathogenic Mutations

To show that this technique could be used to identify sequence variants, we amplified 5′ MR and 5′ LR from genomic DNA isolated from members of 35 pedigrees selected because they had individuals with either very-early-onset disease (VEO), aneurysms, or both. Criteria for VEO disease have been described elsewhere (Fick et al. 1993). Intron-based primers were designed for SSCA screening of exons 11–21 (table 1). Exon 15 is ~3.5 kb in length and was therefore amplified in 18 overlapping fragments.

Using SSCA, we identified eight distinct pathogenic mutations among 10 families in exons 11–21, for a detection rate of ~28.5% (table 2). Seven mutations involved insertions, deletions, or nonsense changes predicted to result in truncated proteins lacking the recepter for egg jelly (REJ) domain, transmembrane spanning regions, and the PKD1 C-terminus. For 6 of the 10 pedigrees, there was more than one family member available for analysis, and the mutation could be shown to segregate with the disease phenotype. One family (UC062) had a C→G transversion at the −3 position of the exon 17 splice-acceptor site. Although, since a cytosine is conserved in this position ~65% of the time, this alteration has a high likelihood of altering the splicing of this exon, we were unable to obtain an RNA sample to demonstrate this effect. The mutation was shared by three affected family members and was not seen in 100 normal chromosomes.

Several families were notable for variability in disease presentation. One example is demonstrated in figure 5. The proband, II-1, of pedigree JHU471 was found to have chronic renal insufficiency in her 50s but was not on dialysis at the time of submission of this article (age ~54 years). Several family members with ADPKD, however, were more severely affected. She had a brother (II-3) who died at age 31 years of a ruptured cerebral aneurysm and a niece (III-1) in whom an asymptomatic intracranial aneurysm was detected that required surgical intervention. Two additional family members (II-4 and III-2) died shortly after birth, as a consequence of VEO disease. The proband was found to have in exon 17 an aberrant SSCA band, which was sequenced and found to contain an 11-bp insertion (fig. 5). Sequence analysis revealed that the insertion contained a local 10-bp duplication, with an additional base pair, a thymidine, added at the beginning of the duplicated sequence. This change was also found in III-1 and II-2. The latter individual is not on dialysis at the age of ~50 years.

Table 2

Mutations Identified in a Panel of 35 Patients with Aneurysms and/or VEO Disease

Pedigree	Phenotype	Mutation	Location	Effect
UC062	VEO	C32, 328G	Intron 16/exon 17	Splice site disrupted
JHU471	VEO + Aneurysm	7397 ins 11	Exon 17	Frameshift
JHU295	Aneurysm	3220 del 4	Exon 13	Frameshift
UC591	VEO + Aneurysm	5224 del 2	Exon 15	Frameshift
UC138	Aneurysm	5224 del 2	Exon 15	Frameshift
UC1334	VEO	5224 del 2	Exon 15	Frameshift
UC125	Aneurysm	5782 del 1	Exon 15	Frameshift
UC834	VEO	3804 del 7	Exon 15	Frameshift
UC109	VEO	G5689A	Exon 15	F1825X
UC362	Aneurysm	6024 ins 1	Exon 15	Frameshift

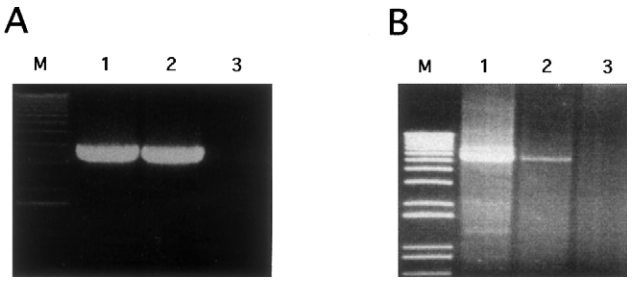

Figure 3 5′ MR (*A*) and 5′ LR (*B*), which are *PKD1* specific. Each template can be amplified from the genomic DNA of an individual (lane 1) and 145.19 (lane 2) but not from the cell line N23HA (lane 3), which contains only the homologous loci. "M" denotes the 1-kb ladder. The "extra" band seen in lane 1 is an artifact of overloading. Hybridization studies of higher-resolution gels, using an internal fragment, have shown that only a single product is amplified with this primer pair.

Three pedigrees, UC138, UC1334, and UC591, were discovered to have an identical 2-bp deletion in exon 15 (table 2 and fig. 6). Family UC138, from the University of Colorado database, was selected for analysis because it contained an individual with a cerebral aneurysm. This individual (III-1; fig. 6) was the only member of the family for whom a DNA sample was available. Family UC1334 was evaluated because of a child (III-1) with VEO disease. The child's affected father (II-1) had the same mutation as was seen in his daughter. The third family, UC591, had several individuals (I-1, II-1, and III-1) with aneurysms, including one (III-1) with both VEO disease and an aneurysm. There were two additional individuals in the family (both cousins of III-1) who had VEO disease but for whom DNA samples were not available. Although each of these three families had individuals who were severely affected, there were also individuals with renal cystic disease and a more routine presentation.

It should be noted that each of these families had been independently ascertained. Each family resides in a different part of the United States (Texas, California, and Minnesota) and is unaware of any distant relatives living in the other communities. Pedigree analysis of at least four generations of each family revealed no genetic relationship between families. In at least two of the cases (pedigrees UC138 and UC1334), the *PKD1* mutation was thought to be from an ancestral white-European source. Both of these families have resided in the United States since at least the 1800s, but in each case there may have been an affected ancestor of Irish descent. Individual I-1 in pedigree 591 had parents of Italian and Native American descent, but it is not known through which parent the disease was transmitted. Since members of family UC591 were informative at a number of intragenic *PKD1* polymorphic loci (described in more detail below; see table 4), we sought to use haplotype analysis to determine whether the mutation in this pedigree was likely to have had an independent origin. We found that the *PKD1* mutation in all of the families occurred on the most common haplotype (i.e., wild-type sequence) for all intragenic *PKD1* polymorphisms (>10 tested). Families 591 and 138 had identical alleles for markers D16S521 (156) and D16S423 (125), but these were also the most frequent. Therefore, we cannot exclude the possibility that the three families share a common, distant genetic origin.

PKD1 Missense Variants Not Found in Normals

Three sequence variants in 5′ MR were found to be absent in 100 normal chromosomes (table 3). Two affected members of pedigree UC1242, a parent and child, were found to have in exon 17 a G→T transition that conserved the valine at position 2378. Although the change is not predicted to affect the protein sequence, it segregates with the disease haplotype. The new codon, GTT, is far less commonly used elsewhere in the gene (19/405 codons for valine) than is the wild-type sequence of GTG (271/405) and thus, theoretically, could affect translational efficiency.

Pedigree UC046 had in exon 17 a G→C transversion that is predicted to result in a nonconservative amino-acid change (Arg→Pro) at position 2392. Although the residue falls within the REJ domain of PKD1, one of the PKD1 portions most highly conserved between humans and *Fugu*, Arg2392 itself is not conserved between the species (Sanford et al. 1997). Thus, we are currently unable to predict the likely consequences of such a change. Unfortunately, only one member of the pedigree was available for evaluation, so we could not determine whether the variant segregated with disease.

Two individuals (parent and child) available from pedigree UC519 were heterozygous for a C→T transition that converts a serine at position 2434 (uncharged polar)

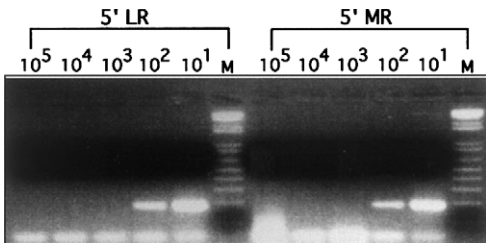

Figure 4 Diluted long-range templates used for amplification of gene-specific fragments. For both 5′ MR (*right panel*) and 5′ LR (*left panel*), dilutions $<10^3$ are not sufficient to remove genomic contamination that is present from the initial PCR reaction, since products outside the template region (exon 32 in this figure) can still be amplified. "M" denotes the 1-kb ladder.

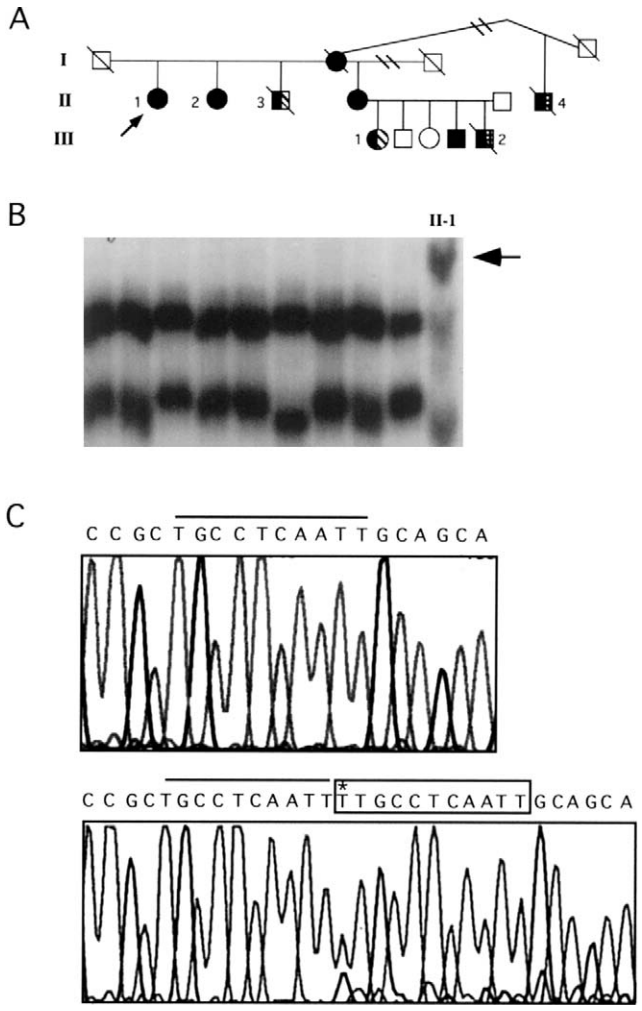

Figure 5 Pedigree JHU471, with an 11-11 insertion in exon 17 and severe disease. A partial pedigree is shown in panel *A*. Members of the family with ADPKD are indicated by blackened symbols. Affected individuals with VEO disease (*hatched symbols*) or cerebral aneurysms (*striped symbols*) are identified. Panel *B* shows the exon 17 SSCA pattern for affected patients. The proband, II-1, has a markedly abnormal pattern. The differences seen among the other patients are due to a common polymorphism in exon 17. Panel *C* shows an aberrant SSCA band (*arrow*) that was excised from the gel and sequenced. The sequence of this band shows an 11-bp insertion, which is indicated by a black line.

to phenylalanine (nonpolar). The amino acid in question is positioned within a highly conserved serine-threonine–rich cluster in the REJ domain, and thus, theoretically, a nonconservative substitution could have pathogenic effects.

Intragenic Polymorphisms

A number of polymorphisms were identified and are listed in table 4. Many of these variants were found to be present in the homologous loci. This was tested by examination of the homologous copies present in the database for the polymorphism in question. If these copies lacked the variant, DNA of the cell line N23HA and total genomic DNA of five individuals (lacking the polymorphism in *PKD1*) were then tested for the presence of the change, by the appropriate restriction digest. On the basis of these studies, we cannot comment on whether homologous loci vary, in terms of sequence, between individuals.

Variants at several of the loci that are physically very close together in the gene appear to be in tight linkage disequilibirum. For example, every white individual (12/12) found to have the exon 20 polymorphism T7919C also had the exon 21 variant A8124G. The same was true for the variants in exons 17 and 18. Of note is that the changes in exons 20 and 21 were found together in one of the *PKD1* homologues in the database, whereas the changes in exons 17 and 18 were not. It is still un-

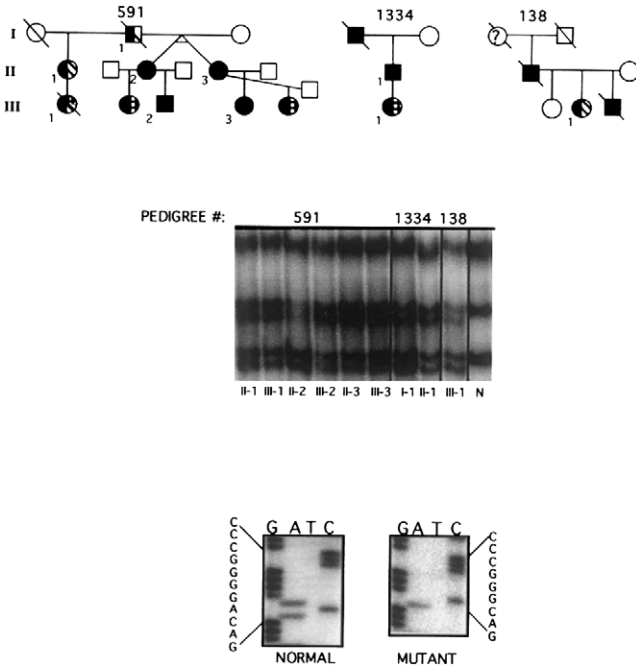

Figure 6 Three independent pedigrees (UC591, UC1334, and UC138) with severe ADPKD and an identical 2-bp deletion in exon 15. The three pedigrees shown in the top panel were ascertained because they had individuals with severe manifestations of ADPKD (*blackened symbols*), including aneurysms (*striped symbols*), and/or VEO disease (*stippled symbols*). The middle panel shows an identical SSCA pattern in exon 15 for members of each pedigree, as indicated. A normal sample (lane N) was included for comparison. The aberrant bands were excised from the gel for each sample and were sequenced (data not shown). The PCR product was also cloned for two individuals (III-1 in pedigree 591, and III-1 in 1334). The bottom panel shows a comparison of the sequences for normal and mutant clones. The mutant clone contains a deletion of 2 bp (AG; table 2) that creates loss of an *Eco*O109I site (data not shown). The pedigrees in the top panel are not complete.

Table 3

Missense and Silent Variants Absent in a Panel of 100 Normal Chromosomes

Pedigree	Phenotype	Exon	Nucleic-Acid Change	Amino-Acid Change	Restriction Site
UC1242	Aneurysm	17	G7345T	Val2378Val	*Csp*6I
UC046	Aneurysm	17	G7386C	Arg2392Pro	*Bsp*120I
UC519	Aneurysm	18	C7479T	Ser2423Phe	*Dsa*I

known whether one of the other homologues whose sequence has not yet been determined contains variants in both exons 17 and 18.

Although most of the variants would not be predicted to alter the protein sequence, two exceptions were discovered. The first was a T→C transition, in exon 15 (T4406C), that leads to the substitution of an arginine residue for a tryptophan at position 1399. Several affected individuals were found to be homozygous for this change, suggesting that it occurs on both affected and unaffected haplotypes. The residue is predicted to lie in position 4 of the B strand of the polycystic kidney–disease (PKD) domain 8 (Bycroft et al. 1999). The substitution does not alter the β-sheet backbone, and arginine is found in the corresponding position of other PKD domains. The other variant, a G→C transversion in exon 19 (G7853C), was noted in a black patient and is predicted to change a glutamate to glutamine at position 2548. Although this difference was not seen in 100 white chromosomes, another black patient (from Belize) was homozygous for the variant. It is unknown whether this substitution would have any biological consequences, since the *Fugu* gene has an arginine in the same position.

Discussion

A number of strategies have been described for mutation detection in the duplicated region of *PKD1*. Most of these methods have taken advantage of unique sequence in the 3′ region of the *PKD1* gene, to design primers that can be used for amplification of long-range PCR products. In general, this has permitted analysis of *PKD1* as far 5′ as exon 15, in cDNA, and exon 23, in genomic DNA. The major limitation has been the length of the PCR products that can be reliably amplified from either cDNA or genomic DNA.

A recent study has described by the use of reverse transcriptase–PCR and *PKD1*-specific primers to amplify the entire *PKD1* coding sequence from RNA (Thongnoppakhun et al. 1999). The method overcomes many of the obstacles encountered by other investigators and can readily identify mutations that affect normal splicing. It has the additional advantage that it can be performed by use of RNA prepared from blood samples. The method has a number of important limitations, however. RNA yields drop rapidly from blood samples that are not processed immediately after phlebotomy. This may preclude the use of samples transported from a distance. In addition, although the use of RNA-based templates offers the possibility of detection of mutations that disrupt splicing, one must assume that the mutant RNA molecules are stable. It is well recognized that some mutations result in RNAs that are rapidly degraded. Finally, one usually seeks to determine the DNA sequence that is responsible for the observed changes in the RNA molecule, to confirm the molecular diagnosis. Therefore, DNA-based techniques are necessary to complement RNA-based strategies for the precise definition of mutations, and vice versa.

We have identified in exon 15 a *PKD1*-unique sequence where both forward and reverse gene-specific primers can be placed. We have used these primers to reliably amplify two templates from genomic DNA that extend the region of analysis to exon 11. Preliminary studies suggest that the reverse primer can be less reliably used with forward primers positioned in more-5′ regions as well (unpublished observations). We have confirmed that these PCR products are indeed gene specific. These reagents should prove useful to investigators who use either DNA or RNA samples for their mutation analyses.

As noted in the introduction, relatively few *PKD1* mutations have been described in the literature, and little correlation has been observed between the pattern of mutation and the phenotype. For the current study, we selected a population of 35 pedigrees with severe manifestations of ADPKD, to evaluate the robustness of our methods. We concentrated on families with either aneurysms or VEO disease, because both features appear to cluster in families. We reasoned that the familial clustering suggested a possible correlation between the genotype and phenotype in these individuals. In 10 of the pedigrees, we identified eight definitive mutations, and all were predicted to result in severely truncated proteins lacking the intracellular C-terminus, the transmembrane-spanning segments, and the REJ domain (if it is assumed that any of them are translated). Although it is tempting to speculate that this group of severely affected patients has more-severe mutations, the data do not yet support this conclusion. The surveyed region contained ~5.3 kb, or 38%, of the coding sequence,

Table 4

Polymorphisms Identified in Exons 11–21

Exon	Nucleic-Acid Change	Amino Acid	Restriction Site	Status in Homologues[a]	Comment
11	G2911A	Pro900Pro	*Sty* I	−	Segregates with C2941T
11	C2941T	Asp910Asp	*Fok* I	−	Segregates with G2911A
13	T3274C	Gly1021Gly	*Eco*O109I	+	
15	A4876C	Ala1555Ala	*Tsp*RI	−	
15	G4885A	Thr1558Thr	*Xcm*I	−	
15	T4406C	Trp1399Arg	*Nci*I	+	
15	C5383T	Thr1724Thr	*Hae*III	−	
17	T7376C	Leu2389Leu	*Mva* II	+	Segregates with C7652T
18	C7652T	Leu2472Leu	*Nsi*I	+	Segregates with T7376C
19	G7853C	Glu2548Gln	*Taq*I	−	
20	T7919C	Leu2570Leu	*Bsa* I	+	Segregates with A8124G
21	A8124G	His2638Arg	*Hha* I	+	Segregates with T7919C

[a] A minus sign (−) denotes absence, and a plus sign (+) denotes presence.

which is consistent with our detection rate (~29%) in the region.

There was one surprising result, however, that does suggest a potential correlation between the severe phenotype and genotype. We have identified three separate pedigrees (~9% of the total evaluated) that have an identical mutation (5224del2) in exon 15 of *PKD1*. Although we cannot exclude the possibility of a distant common genetic origin, it is important to emphasize that the families were prospectively selected for evaluation solely on the basis of their phenotype. Other groups of investigators failed to detect this mutation in either 135 affected individuals of European ancestry, by use of the protein-truncation test (Roelfsma et al. 1997), or 24 unrelated individuals screened by direct sequence analysis (Thomas et al. 1999). These observations suggest that the 5224del2 mutation is not a common cause of PKD in populations of European ancestry. Thus, the likelihood that we would have randomly discovered three families sharing this uncommon mutation is low ($P < 0.008$; Fisher's t test).

Further evidence suggesting a possible relationship between the genotype and phenotype was recently presented by Hateboer et al. (1999). These investigators compared the clinical profiles of 10 large Welsh families, each of which had *PKD1*-linked disease with a unique disease-associated haplotype. Interfamilial differences were observed in a number of clinical parameters, including renal survival. The authors concluded that phenotypic differences exist between families and that these are most likely due to the nature of the underling *PKD1* mutation.

Although both sets of observations are provocative, there is ample evidence that refutes a straightforward correlation between a particular genotype and a more severe phenotype; for example, mutations in diverse segments of *PKD1* have been associated with more severe clinical presentations (Peral et al. 1996; Roelfsma et al. 1997). Furthermore, one might expect that affected individuals within a family would have similar disease severity. Although two of the three families described above had more than one severely affected member, all three families also had members with a more classic clinical presentation. A particularly striking example of discordant phenotypes has been described by Peral et al. (1996). One of a pair of DZ twins presented with VEO and a stable nonsense mutation near the 3′ end of the gene (Tyr3818Stop), whereas the other twin had the mutation but no evidence of cysts, consistent with an adult-onset disease course.

There are several ways to reconcile the familial clustering of severe manifestations with the contradictory observations enumerated above. First, several studies have suggested that the rate at which somatic mutations accumulate is likely to be one of the major factors that determine disease severity (Qian et al. 1996; Brasier and Henske 1997; Watnick et al. 1998*b*; Koptides et al. 1999). It is possible that certain mutant haplotypes may influence the rate at which second hits occur, possibly by promoting mitotic recombination between normal and mutant alleles (gene conversion). It is also conceivable that the rate of "second hits" may be further influenced by characteristics of the "normal" allele that have yet to be defined. This could contribute to the phenotypic differences that are commonly observed between family members with the same germline mutation. Last, certain mutations (regardless of their location) may result in mutant gene products with distinct biological properties that may, for example, predispose to aneurysm formation. What seems clear from these considerations is that the factors involved in determination of the familial clustering of certain severe manifestations of ADPKD are likely to be complex. Further directed study of mutant (and "normal") alleles identified in these families is war-

ranted and may help to define the basis for more-severe phenotypes.

In this report, we also have described 10 novel intragenic polymorphisms that have several noteworthy features. A number of them (table 4) could be demonstrated to be present in the homologous loci, supporting previous data that these homologues may serve as a reservoir for sequence variants via gene-conversion events (Watnick et al. 1998*a*). The observation that several sets of alleles appear to be in tight linkage disequilibrium suggests that each group of changes could have been introduced into the *PKD1* gene as the result of a single ancient gene-conversion event. Not all of the polymorphisms, however, were found to be present in the homologues; for example, two changes (G2911A and C2941T) that were physically and genetically linked in exon 11 could not be shown to be present in either N23HA or the blood DNA of five additional individuals. It is still possible, however, that these changes arose as the result of a gene-conversion event, since the sequence of the homologous loci may not be uniform between individuals. The existence of gene-conversion events makes it critical that rigorously tested, locus-specific reagents be used for mutation detection.

Although truncating mutations of *PKD1* are generally accepted to be pathogenic, our results highlight the difficulties in the analysis of amino-acid substitutions in this protein. One black individual, for example, had in exon 19 a nonconservative amino-acid change that was not found in ~80 whites. It was excluded as a pathogenic change, since another individual (also black) was homozygous for the same variant. Not surprisingly, therefore, with respect to the normal *PKD1* sequence, there are likely to be interracial and interethnic differences that will need to be verified by use of ethnically matched control populations.

Two additional nonconservative amino-acid changes, absent from a control population, were also identified. Even though one of these changes segregated with the affected haplotype, it is difficult to conclude that this isolated change is pathogenic. One interesting hypothesis is that in some cases, rather than there being a single pathogenic mutation, the combined effect of a number of missense substitutions results in a hypomorphic allele. As more understanding is gained with respect to the biology of *PKD1*, it should be possible to examine this hypothesis by the testing of missense variants in functional assays. This type of analysis may also give important clues regarding critical areas of the protein structure.

Acknowledgments

We thank the many individuals and families who have participated in these studies. We also gratefully acknowledge the assistance of Dr. Klaus Piontek and Ms. Sidney McGaugey in the preparation of the manuscript. Elizabeth Lyden assisted with the statistical analysis. This work was supported by the National Institutes of Health grants DK 48006, TW05393, DK4853, and 5T01DK34039 and by the National Kidney Foundation of Maryland. G.G.G. is the Irving Blum Scholar of the Johns Hopkins University School of Medicine.

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

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

Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff, http://www.uwcm.ac.uk/uwcm/mg/hgmd0.html

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for ADPKD [MIM173900] and PKD1 [MIM601313])

References

American PKD1 Consortium (1995) Analysis of the genomic sequence for the autosomal dominant polycystic kidney disease (PKD1) gene predicts the presence of a leucine-rich repeat. Hum Mol Genet 4:575–582

Brasier RL, Henske EP (1997) Loss of the polycystic kidney disease (PKD1) region of chromosome 16p13 in renal cyst cells supports a loss of function model for cyst pathogenesis. J Clin Invest 99:194–199

Brook-Carter PT, Peral B, Ward CJ, Thompson P, Hughes J, Maheshwar MM, Nellist M, et al (1994) Deletion of the TSC2 and PKD1 genes associated with severe infantile polycystic kidney disease—a contiguous gene syndrome. Nat Genet 8:328–332

Bycroft M, Bateman A, Clark J, Hamill SJ, Sanford 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

Ceccherini I, Persici P, Pezzolo A, Rocchi M, Breuning MH, Himmelbauer H, Frischauf A-M, et al (1992) Construction of a fine structure map of chromosome 16 by using radiation hybrids. Proc Natl Acad Sci USA 89:104–108

Chapman AB, Rubinstein D, Hughes R, Stears JC, Earnest MP, Johnson AM, Gabow PA, et al (1992) Intracranial aneurysms in autosomal dominant polycystic kidney disease. New Engl J Med 327:916–920

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

Fick GM, Johnson AM, Strain JD, Kimberling WJ, Kumar S, Manco-Johnson ML, Duley IT, et al (1993) Characteristics of very early onset autosomal dominant polycystic kidney disease. J Am Soc Nephrol 3:1863–1870

Gabow PA, Kaehny WD, Johnson AM, Duley IT, Manco-John-

son M, Lezotte DC, Schrier RW (1989) The clinical utility of renal concentrating capacity in polycystic kidney disease. Kidney Int 35:675–680

Germino GG, Barton NJ, Lamb J, Higgs DR, Harris P, Scherer G, Nakamura Y, et al (1990) Identification of a locus which shows no genetic recombination with the autosomal dominant polycystic kidney disease gene on chromosome 16. Am J Hum Genet 46:925–933

Germino GG, Weinstat-Saslow D, Himmelbauer H, Gillespie GAJ, Somlo S, Wirth B, Barton N, et al (1992) The gene for autosomal dominant polycystic kidney disease lies in a 750-kb CpG-rich region. Genomics 13:144–151

Hateboer N, Lazarou LP, Williams AJ, Holmans P, Ravine D (1999) Familial phenotype differences in PKD1. Kidney Int 56:34–40

Hughes J, Ward CJ, Peral B, Aspinwall R, Clark K, San Millan 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

International Polycystic Kidney Disease Consortium (1995) Polycystic kidney disease: the complete structure of the PKD1 gene and its protein. Cell 81:289–298

Koptides M, Hadjimichael C, Koupepidou P, Pierides A, Deltas CC (1999) Germinal and somatic mutations in the PKD2 gene of renal cysts in autosomal dominant polycystic kidney disease. Hum Mol Genet 8:509–513

Peral B, Gamble V, Strong C, Ong ACM, Sloane-Stanley J, Zerres K, Winearls C, 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 stable, nonsense mutation associated with a case of infantile onset polycystic kidney disease 1 (PKD1). Hum Mol Genet 5:539–542

Qian F, Watnick TJ, Onuchic LF, Germino GG (1996) The molecular basis of focal cyst formation in human autosomal dominant polycystic kidney disease type 1. Cell 87:979–987

Roelfsema, JH, Spruit L, Saris JJ, Chang P, Pirson Y, van Ommen GJB, 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

Shoffner JM (1998) Molecular analysis of oxidative phosphorylation diseases for detection of mitochondrial DNA mutations. In: Dracopoli NC, Haines JL, Korf BR, Moir DT, Morton CC, Seidman CE, Seidman JG, et al (eds) Current protocols in human genetics 9, clinical molecular genetics. John Wiley & Sons, New York, pp 9.9.1–9.9.26

Thomas R, McConnell R, Whittacker J, Kirkpatrick P, Bradley J, Sandford R (1999) Identification of mutations in the repeated part of the autosomal dominant polycystic kidney disease type 1 gene, PKD1, by long-range PCR. Am J Hum Genet 65:39–49

Thongnoppakhun W, Wilairat P, Kriengsak V, Yenchitsomanus P (1999) Long RT-PCR amplification of the entire coding sequence of the polycystic kidney disease 1 (PKD1) gene. Biotechniques 26:126–132

Van Raay TJ, Burn TC, Connors TD, Petry LR, Germino GG, Klinger KW, Landes GM (1996) A 2.5kb polypyrimidine tract in the PKD1 gene contains at least 23 H-DNA-forming sequences. Microb Comp Genet 1:317–327

Watnick TJ, Gandolph MA, Weber H, Neumann HPH, 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

Zerres K, Rudnik-Schoneborn S, Deget F, members of the German working group on paediatric nephrology (1993) Childhood onset autosomal dominant polycystic kidney diseas in sibs: clinical picture and recurrence risk. J Med Genet 30:583–588