

1-1-2002

## **Mutation detection for genotype/phenotype correlation studies in autosomal dominant polycystic kidney disease**

Marie McCluskey  
*Edith Cowan University*

Follow this and additional works at: <https://ro.ecu.edu.au/theses>



Part of the [Medicine and Health Sciences Commons](#)

---

### **Recommended Citation**

McCluskey, M. (2002). *Mutation detection for genotype/phenotype correlation studies in autosomal dominant polycystic kidney disease*. Edith Cowan University. Retrieved from <https://ro.ecu.edu.au/theses/723>

This Thesis is posted at Research Online.  
<https://ro.ecu.edu.au/theses/723>

# Edith Cowan University

## Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study.

The University does not authorize you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following:

- Copyright owners are entitled to take legal action against persons who infringe their copyright.
- A reproduction of material that is protected by copyright may be a copyright infringement. Where the reproduction of such material is done without attribution of authorship, with false attribution of authorship or the authorship is treated in a derogatory manner, this may be a breach of the author's moral rights contained in Part IX of the Copyright Act 1968 (Cth).
- Courts have the power to impose a wide range of civil and criminal sanctions for infringement of copyright, infringement of moral rights and other offences under the Copyright Act 1968 (Cth). Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

**Mutation detection for genotype/phenotype correlation studies in  
autosomal dominant polycystic kidney disease**

**by**

**Marie McCluskey B.App.Sci. (Hons)**

**A Thesis Submitted for the Award of  
Doctor of Philosophy in Human Biology**

**At the Faculty of Communications, Health and Science,  
Edith Cowan University, Joondalup.**

**Date of submission: April 15<sup>th</sup>, 2002**

## ABSTRACT

Autosomal dominant polycystic kidney disease (ADPKD) is one of the most common genetic disorders (Gabow, 1993), with an estimated prevalence of 1 in 400 to 1 in 1000 (Dalgaard, 1957; Gabow, 1993). The disease is the fourth leading cause of renal failure with more than 10 million people affected worldwide (Grantham, 1997). Mutations in at least two genes (*PKD1*, *PKD2*) can lead to the disease (European PKD Consortium, 1994; Mochizuki, *et al.* 1996). This project deals with autosomal dominant PKD, caused by mutations in the *PKD1* gene, which account for 85% of reported cases.

Mutation detection in the *PKD1* gene is difficult as the gene is relatively large (46 exons over 52kb of genomic DNA), it has a high GC content, and is duplicated in numerous homologues of unknown function. Mutations detected to date are usually private and many neutral polymorphisms have also been detected. Co-amplification of *PKD1* homologous genes, which correspond to at least 75% of its mRNA, poses the largest obstacle for *PKD1* mutation detection studies.

In this study, novel strategies for mutation screening in the duplicated region of *PKD1* have been developed. In the process of the work further characterisation of the *PKD1* homologous genes has been achieved, and the results suggest that at least two of the homologues are non-functioning (*i.e.* not translated) pseudogenes.

Subjects from three countries were screened for mutations in the *PKD1* gene in order to detect mutations and to attempt genotype-phenotype correlations in a disorder which



exhibits a high degree of inter- and intra-familial phenotypic heterogeneity. Seventy-four sequence variations were detected, of which twenty-four are believed to be pathogenic.

In the Australian cohort, DNA was screened over 95% of the coding region and a mutation detection rate of 74% was achieved. This result is favourable when compared to other studies. Findings from the study and others in the literature suggest that *PKDI* germline mutations do not indicate a simple genotype-phenotype correlation, and future studies will need to incorporate many other biological and environmental variables in order to establish how the genotype influences the complex phenotype of the disease.

## DECLARATION

I certify that this thesis does not, to the best of my knowledge and belief:

- (i) incorporate without acknowledgment any material previously submitted for a degree or diploma in any institution of higher education;
- (ii) contain any material previously published or written by another person except where due reference is made in the text; or
- (iii) contain any defamatory material.

Signature

Date 14/10/02.

## ACKNOWLEDGEMENTS

I wish to thank my supervisors, Assoc. Prof. Luba Kalaydjieva and Prof. Alan Bittles at Edith Cowan University, Joondalup, WA and Dr. Steve Wilton at the Australian Neuromuscular Research Institute (ANRI), Nedlands, WA. I would also like to thank my additional supervisors Dr. Bernd Dworniczak and Prof. Jürgen Horst at the Institute for Human Genetics, Münster, Germany.

This project would not have been possible without the ADPKD patients and families who participated in Australia, Bulgaria and Poland. I would like to offer my sincere thanks to them and their clinicians, particularly Dr. Mark Thomas, Dr Steve Melsom and Ken Roberts at Royal Perth Hospital, Perth, WA; Dr V. Todorov, (Plovdiv) & Dr. D. Dimitrakov (Pleven) Bulgaria; and Dr. Peter Jasik, University Of Krakow, Poland.

Thanks to my colleagues for their contributions: Tina Schiavello, Dave Chandler, Dora Angelicheva, Bec Gooding, Wei Wang, Michael Hunter and Janina Hantke in Perth at ECU; Lori Blechynden, Kaite Honeyman, Sue Fletcher and Caroline O'Leary at ANRI, WA; and in Germany, Nadja Bogdanova, Arseni Markoff, Petra Pennekamp, Karin Sickmann, Karl Hackmann, Susi Brandt, Sabine Preisler-Adams and Jürgen Wansch.

If I have unintentionally overlooked any participants I apologise. In a large multi-centre study of this nature it is difficult to name you all, please don't think that you are forgotten, I am grateful to all of you.

I am extremely grateful to both Edith Cowan University, and IKF (NRW, Germany) who awarded scholarships to me during my candidature.

Thanks to my friends (you know who you are) for your support in this endeavour. To my family, I want you to know that you are all sincerely appreciated. Special thanks go to my parents who have provided support and always believed in me, even when I didn't believe in myself.

## DEFINITION OF ABBREVIATIONS AND TERMS

ADPKD	Autosomal dominant polycystic kidney disease
ARPKD	Autosomal recessive polycystic kidney disease
bp	Nucleotide base pairs
cDNA	Complementary deoxyribonucleic acid
CMC	Chemical mismatch cleavage
CRF	Chronic renal failure
CSGE	Conformational sensitive gel electrophoresis
DGGE	Denaturing gradient gel electrophoresis
DHPLC	Denaturing high performance liquid chromatography
DNA	Deoxyribonucleic acid
EMC	Enzymatic mismatch cleavage
Endo VII	T4 endonuclease VII
ESRF	End stage renal failure
FN-III	Fibronectin type-3 repeat
HDA	Hetero Duplex Analysis
HG(s)	Homologous gene or genes
3'-HVR	Three prime hypervariable repeat region of the alpha-globin gene.
IVS	Intervening Sequence (intronic region)
kb	Kilobase (equivalent to 1000 bp)
LDL-A	Low density lipoprotein type-A
LRR	Leucine rich repeat

mRNA	Messenger ribonucleic acid
NIRCA	Non-isotopic RNase Cleavage assay
nt	Nucleotide
PAH	Phenylalanine hydroxylase gene
PBP	Polycystic breakpoint
PCR	Polymerase chain reaction
PKD	Polycystic kidney disease
<i>PKD1/2/3</i>	Polycystic kidney disease gene (1, 2 or 3).
<i>PKD1P1-6</i>	Polycystic kidney disease gene 1 pseudogene (1-6)
PKU	Phenylketonuria
PTT	Protein truncation test
REJ	Receptor egg jelly domain/module
RT-PCR	Reverse transcription-Polymerase chain reaction
SSCP	Single-strand conformation polymorphism
TSC	Tuberous sclerosis
<i>TSC2</i>	Tuberous sclerosis 2 gene

## LIST OF TABLES

	<b>Page</b>
<b>Table II.1</b> Diseases Reported to Co-occur with ADPKD	11
<b>Table II.2</b> <i>PKD1</i> Mutations by Type and Region	37
<b>Table II.3</b> Mutations Detected in <i>PKD1</i>	43-51
<b>Table II.4(a)</b> Non-disease Causing Missense Changes Detected in <i>PKD1</i>	55
<b>Table II.4(b)</b> Neutral Polymorphisms Detected in <i>PKD1</i>	56-58
<b>Table II.5</b> Phenotype associated with Q4041X	64
 <b>Table III.1</b> Mutation Screening of DNA in Exons 34 to 46 of the <i>PKD1</i> Gene	 86
<b>Table III.2</b> Mutations Detected in the Duplicated Region of the <i>PKD1</i> gene	87
<b>Table III.3</b> PCR Amplification Reagents	95
<b>Table III.4</b> PCR Cycling Steps	96
<b>Table III.5</b> PCR of <i>PKD1</i> Unique region	97
<b>Table III.6</b> Exon 42 Amplification Primers	98
<b>Table III.7</b> Exon-Specific SSCP Conditions	100
<b>Table III.8</b> Asymmetric PCR Amplification Reagents	103
<b>Table III.9</b> Isotopic Labelling Reaction	104
<b>Table III.10</b> Sequencing of <i>PKD1</i> Unique region	105
<b>Table III.11</b> Termination Mixes for Sequencing	106
<b>Table III.12</b> Automated Sequencing Protocol	109
<b>Table III.13</b> Primers for DNA Pool Screening	115

<b>Table III.14</b>	Primers for BAC and PAC DNA Amplification	116
<b>Table III.15</b>	Primers for XL-Template Amplification	118
<b>Table III.16</b>	Primers for the PCR Amplification of <i>PKD1</i> Exons 2-5	120
<b>Table III.17</b>	Primers for the PCR Amplification of <i>PKD1</i> Exons 6-11	121
<b>Table III.18</b>	Primers for the PCR Amplification of <i>PKD1</i> Exons 12-14	121
<b>Table III.19</b>	Primers for the PCR Amplification of <i>PKD1</i> Exon 15	122
<b>Table III.20</b>	Primers for the PCR Amplification of <i>PKD1</i> Exons 15-21	123
<b>Table III.21</b>	Primers for the PCR Amplification of <i>PKD1</i> Exons 23-34	124
<b>Table III.22</b>	Primers for Amplification of <i>PKD1</i> and Homologous Genes	128
<b>Table V.1</b>	Summary of probable disease-causing mutations in <i>PKD1</i>	145-6
<b>Table VI.1</b>	Silent mutations in the <i>PKD1</i> gene	209-12
<b>Table VI.2</b>	Missense variations in <i>PKD1</i> not segregating with the disease	213
<b>Table VI.3</b>	Shared polymorphisms of the 5' Duplicated region of <i>PKD1</i>	242
<b>Table VI.4</b>	The shared common polymorphisms of <i>PKD1</i> exons 44-46	244
<b>Table VII.1</b>	Efficiency of <i>PKD1</i> screening	252
<b>Table VII.2</b>	Observed Lod score values for <i>PKD1</i> families (mutation undetected)	253
<b>Table VII.3</b>	Polymorphisms in <i>PKD1</i> -linked patients with undetected mutation	268
<b>Table VII.4</b>	Shared pathogenic mutations in <i>PKD1</i>	272
<b>Table VII.5</b>	Private mutations with a severe ADPKD phenotype	273
<b>Table A1</b>	Enzyme Buffers for Optimisation of $Mg^{2+}$ Concentrations	312

## LIST OF FIGURES

	<b>Page</b>
<b>Fig II.1</b>	The <i>PKD1</i> Gene 22
<b>Fig II.2</b>	Polycystin-1 23
<b>Fig II.3</b>	Mutations of <i>PKD1</i> 40
 <b>Fig III.1</b>	 Autoradiograph showing wild-type and mutant sequences 108
 <b>Fig IV.1</b>	 Amplification results of BAC and PAC clones 132
<b>Fig IV.2</b>	XL-PCR primers used for the amplification of <i>PKD1</i> (Exons 2-33) 136
<b>Fig IV.3</b>	Alignment of <i>PKD1</i> & HG sequences in the IVS11 Primer Region 137
<b>Fig IV.4</b>	Co-amplification of PKD1 & HGs in nested PCR 138
 <b>Fig V.1</b>	 Results of XL amplification in the duplicated region of <i>PKD1</i> 141
<b>Fig V.2</b>	SSCP optimisation for mutation screening 142
<b>Fig V.3</b>	Proposed effect of splice site mutation g.48491C>A 180
<b>Fig V.4</b>	Evolutionary conservation & topology prediction for G381C 184
<b>Fig V.5</b>	Evolutionary conservation & topology prediction for Y2185D 186
<b>Fig V.6</b>	Evolutionary conservation & topology prediction for G2785D 189
<b>Fig V.7</b>	Evolutionary conservation & topology prediction for L4046F 191
<b>Fig V.8</b>	Evolutionary conservation & topology prediction for R4227P 194
<b>Fig V.9</b>	Evolutionary conservation & topology prediction for c.7472del13 197
<b>Fig V.10</b>	Evolutionary conservation & topology prediction for c.9292del39 199



<b>Fig V.11</b>	Evolutionary conservation & topology prediction for c.l1669del6	202
<b>Fig V.12</b>	Nucleotide insertion & corresponding amino acids (c.l3069ins39)	204
<b>Fig V.13</b>	Evolutionary conservation & topology prediction for c.l3069ins39	205
<b>Fig VI.1</b>	Evolutionary conservation & topology prediction for Q739R	215
<b>Fig VI.2</b>	Evolutionary conservation & topology prediction for M1092T	216
<b>Fig VI.3</b>	Evolutionary conservation & topology prediction for W1399R	218
<b>Fig VI.4</b>	Evolutionary conservation & topology prediction for T1649M	220
<b>Fig VI.5</b>	Evolutionary conservation & topology prediction for H2638R	222
<b>Fig VI.6</b>	Evolutionary conservation & topology prediction for R2761C	224
<b>Fig VI.7</b>	Evolutionary conservation & topology prediction for F3066L	226
<b>Fig VI.8</b>	Evolutionary conservation & topology prediction for T3509M	228
<b>Fig VI.9</b>	Evolutionary conservation & topology prediction for S3592I	230
<b>Fig VI.10</b>	Evolutionary conservation & topology prediction for G3695Q	232
<b>Fig VI.11</b>	Evolutionary conservation & topology prediction for S3791T	234
<b>Fig VI.12</b>	Evolutionary conservation & topology prediction for I4044V	236
<b>Fig VI.13</b>	Evolutionary conservation & topology prediction for A4058V	238
<b>Fig VI.14</b>	B23.1 Clones, wild-type allele & the IVS41 +5insggg	240

## TABLE OF CONTENTS

	Page(s)
<b>I. Introduction</b>	<b>1-5</b>
<b><u>I.1. Introduction</u></b>	<b>1-3</b>
1.1.i. Autosomal Dominant Polycystic Kidney Disease	2
1.1.ii. The <i>PKD1</i> Gene	2-3
1.1.iii. The ADPKD Phenotype	3
<b><u>I.2 Strategy of the Study</u></b>	<b>4-5</b>
1.2.i. Study Aims & Expected Outcomes	4
1.2.ii. Study Plan	4-5
 <b>II. Review of Literature</b>	 <b>6-70</b>
<b><u>II.1. General Literature</u></b>	<b>7-15</b>
II.1.i. <u>Clinical Course and Heterogeneity of ADPKD</u>	7-11
II.1.ii. <u>The Pathogenesis of ADPKD</u>	12-15
<b><u>II.2. The Genetic Basis of ADPKD</u></b>	<b>16-32</b>
II.2.i. <u>Locus Heterogeneity in ADPKD</u>	16-20
II.2.ii. <u>Ethnic Heterogeneity of ADPKD</u>	20
II.2.iii. <u>Molecular Cloning of the PKD Genes</u>	20-21
II.2.iv. <u>Structure of the <i>PKD1</i> Gene</u>	22
II.2.v. <u>The <i>PKD1</i> Protein</u>	22-27
II.2.vi. <u><i>PKD1</i> Homologous Transcripts</u>	28
II.2.vii. <u>Polycystic Kidney Disease in Animals</u>	28-29

II.2.viii. <u>The Structure of the <i>PKD2</i> Gene</u>	29-30
II.2.ix. <u>Expression Patterns of Polycystin-1 and Polycystin-2</u>	30-32
II.3. <u>The Mutations and Polymorphisms Associated with ADPKD</u>	33-58
II.3.i. <u>Mutational Mechanisms of ADPKD</u>	33-36
II.3.ii. <u>Mutations Found in <i>PKD1</i></u>	37
II.3.iii. <u>Types of Mutations in <i>PKD1</i></u>	37
II.3.iv. <u>Location of <i>PKD1</i> Mutations</u>	38-40
II.3.v. <u>Recurrent Mutations in <i>PKD1</i></u>	41
II.3.vi. <u>Mutational Effects on Protein</u>	41-51
II.3.vii. <u>Neutral Polymorphisms in the Human Genome</u>	52-53
II.3.viii. <u>Neutral Polymorphisms in <i>PKD1</i></u>	53-58
II.4. <u>Genotype-Phenotype Correlation in Single-Gene Disorders</u>	59-70
II.4.i. <u>Phenotypic Variation</u>	60
II.4.ii. <u>Genotype-Phenotype Correlation Studies</u>	60-61
II.4.ii.(a). <u>Locus Heterogeneity</u>	61-64
II.4.ii.(b). <u>Mutation Heterogeneity</u>	65
II.4.ii.(c). <u>Compound Heterozygotes and Double Mutant Alleles</u>	65
II.4.ii.(d). <u>Neutral Variants</u>	66-67
II.4.ii.(e). <u>Modifying Genes</u>	67-68
II.4.iii. <u>Environmental Factors</u>	68-69
II.4.vii. <u>The importance of Genotype-Phenotype Studies for ADPKD</u>	69-70

<b>III. Methodology</b>	<b>71-128</b>
<b><u>III. Methodology (Part I)</u></b>	<b>72-90</b>
III.1 Literature Review of Methodology	72-90
III.1.i. Pre-Screening Methods	73
III.1.ii. DNA Screening Techniques.	73-80
III.1.ii.(a). Single-Strand Conformation Polymorphism	
Analysis	73-75
III.1.ii.(b). HDA, CSGE, DGGE and DHPLC.	75-77
III.1.ii.(c). Dideoxyfingerprinting	77-78
III.1.ii.(d). Mismatch Cleavage Methods	78
III.1.ii.(e). DNA Sequencing	79-80
III.1.iii. RNA Screening Techniques.	80
III.1.iv. Protein Screening Techniques	81-82
III.1.v. Comparison of Mutation Detection Methods	82-83
III.1.vi. Amplification of <i>PKD1</i> without Homologous Gene (HG)	
Contamination	84-85
III.1.vii. Methodology specific to ADPKD	85-90
<b><u>III. Methodology (Part II)</u></b>	<b>91-128</b>
<u>III.2-7. Subjects and Methods</u>	91-128
<u>III.2. Additional Protocols</u>	91



III.5. Mutation Detection in the 5' Duplicated Region	113-119
III.5.i. Amplification of <i>PKDI</i> -specific template in the 5' Duplicated Region	113-114
III.5.ii. Screening of BAC and PAC DNA Pools for <i>PKDI</i> , and the HGs	114
III.5.ii(a). Screening BAC and PAC Pools	114-115
III.5.ii.(b). Culture and Isolation of BAC and PAC DNA	116
III.5.ii.(c). Amplification and Sequencing of BAC and PAC DNA	116-117
III.5.ii.(d). <i>PKDI</i> -Specific Sequence in the Duplicated Region of the Gene	117-119
III.6. Mutation Detection Screen of <i>PKDI</i> Exons 2 to 33	119-125
III.6.i. PCR Amplification of individual <i>PKDI</i> Exons	120
III.6.i.(a) Amplification of exons 2 to 5	120
III.6.i.(b) Amplification of exons 6 to 11	121
III.6.i.(c) Amplification of exons 12 to 14	121
III.6.i.(d) Amplification of exon 15	122
III.6.i.(e) Amplification of exons 15 to 21	123
III.6.i.(f) Amplification of exons 23 to 33	124
III.6.ii. Purification and Sequencing of Nested PCR Products	125
III.7. Investigation of <i>PKDI</i> Homologous Gene Sequences	125-128
III.7.i. Reverse Transcription and PCR (RT-PCR)	125-128

<b>IV. Results –</b>	
<b>Specific Amplification of the Duplicated Region of PKD1</b>	<b>129-139</b>
IV.1. <u>Introduction</u>	130
IV.2. Identification of <u>human</u> genomic clones containing <i>PKD1</i> , and its homologous genes	131-132
IV.3. Sequence analysis of <i>PKD1</i> and the homologous genes	132-134
IV.4. Functional characteristics of the homologous genes	134-135
IV.5. <u>Design</u> of PKD1-Specific Oligonucleotides	135-137
IV.6. <u>Specificity</u> of the PCR amplification	138-139
 <b>V. Results – Mutations</b>	 <b>140-205</b>
Chapter V Introduction	141-146
Truncating Mutations	147-174
Splice-Site Variation	175-181
Missense Variation	182-194
Non-Truncating Insertions & Deletions	195-205

<b>VI. Results – Polymorphisms</b>	<b>206-244</b>
VI.1. Introduction	207-208
VI.2. Missense Polymorphisms	208-239
VI.3. A variant in the 3' Unique Region segregating with the disease phenotype	240-241
VI.4. Linkage Disequilibrium in the <i>PKD1</i> gene	241-244
VI.4.i. Linkage Disequilibrium in the 5' Duplicated Region of <i>PKD1</i>	241-242
VI.4.ii. Linkage Disequilibrium in the 3' Unique Region of <i>PKD1</i>	243-244
<b>VII. Discussion</b>	<b>245-276</b>
VII. Discussion	246-273
Conclusions and Future Studies	274-276
<b>VIII. Bibliography</b>	<b>277-310</b>



<b>Appendices</b>	<b>311-318</b>
<b>Appendix A. List of Protocols from Methods Chapter</b>	<b>311-318</b>
A1. PCR <u>O</u> ptimisation with Differing buffers	312
A2. <u>Q</u> IAquick PCR Purification Protocol	312-313
A3. Centricon Filter Protocol	313
A4. Cleaning ABI sequencing reactions prior to electrophoresis	313-314
A5. Plasmid Purification Post BAC/PAC clone identification	314-315
A6. ABI 373 DNA <u>S</u> equencing (Perkin Elmer)	316-317
A7. Sequitherm <u>E</u> xcel II Cycle Sequencing	317-318
 Appendix B. <u>J</u> ournal articles resulting directly from this thesis (enclosed)	 318
 B1. Bogdanova & McCluskey <i>et al.</i> (2000). Screening the 3' region of the polycystic kidney disease 1 (PKD1) gene in 41 Bulgarian and Australian kindreds reveals a prevalence of protein truncation mutations. <i>Hum Mutat.</i> 16(2): 166-174.	
 B2. Bogdanova <i>et al.</i> (2001). Homologues to the first gene for polycystic kidney disease are pseudogenes. <i>Genomics.</i> 74(3): 333-341.	
 B3. McCluskey <i>et al.</i> (2002). Mutation detection in the duplicated region of the polycystic kidney disease 1 (PKD1) gene in PKD1-linked Australian families. <i>Hum Mutat.</i> 19(3): 240-250.	

## **I. Introduction**

# **I. Introduction**

## **I.1. Introduction**

### **I.1.i. Autosomal Dominant Polycystic Kidney Disease**

Autosomal dominant polycystic kidney disease (ADPKD) is one of the most common genetic disorders (Gabow, 1993), with an estimated prevalence of 1 in 400 to 1 in 1000 (Dalgaard, 1957; Gabow, 1993). The disease is the fourth leading cause of renal failure with more than 10 million people affected worldwide (Grantham, 1997).

Although most cases of the disease present in adulthood, hence the synonym adult polycystic kidney disease, it is in fact a systemic hereditary disorder which can occur at any time of life, including *in utero* (Zerres *et al.*, 1985). The mode of inheritance in ADPKD was defined over 40 years ago and the disease exhibits complete penetrance with respect to the occurrence of renal cysts (Dalgaard, 1957), although it does not always progress to chronic or end-stage renal failure (Churchill, *et al.*, 1984). Mutations in at least two genes (*PKD1*, *PKD2*) can lead to the disease (European PKD Consortium, 1994; Mochizuki, *et al.* 1996). This project deals with autosomal dominant PKD, caused by mutations in the *PKD1* gene.

### **I.1.ii. The *PKD1* Gene**

The first gene for polycystic kidney disease (*PKD1*) is located on human chromosome 16p13.3 (The European PKD Consortium, 1994). The gene produces a 14 kb transcript

which has been detected in many tissue-specific cell lines (The European PKD Consortium, 1994). The highest *PKD1* gene expression levels were found in kidney- and liver-derived cell lines (The European PKD Consortium, 1994). The *PKD1* gene spans 52 kb of genomic DNA, and includes 46 exons. A 14 kb mRNA transcript is produced and encodes a novel protein, polycystin-1, which is 4302 amino acids in length (The International PKD Consortium, 1995). *PKD1* is duplicated at least 3 times on human chromosome 16p13.1. The *PKD1* homologous genes (HGs) are up to 97% homologous with the *PKD1* gene over two-thirds of its length (The European PKD Consortium, 1994). It is not known if the HG transcripts produce functional proteins (Hughes *et al.*, 1995). Due to their co-amplification with *PKD1*, they represent a major obstacle for mutation detection studies in the *PKD1* gene (Peral *et al.*, 1997).

#### I.1.iii. The ADPKD Phenotype

Both inter- and intra-familial phenotypic variation have been observed in autosomal dominant polycystic kidney disease. Some affected individuals do not progress to end-stage renal failure (ESRF), and those that do exhibit a wide variation in age at onset of ESRF (Milutinovic *et al.*, 1984; Parfrey *et al.*, 1990; Gabow, 1993; Grünfeld, 1998). In addition to renal disease, ADPKD has been associated with a number of extra-renal symptoms, these also vary both between and within affected families (Gabow, 1993).

## I.2 Strategy of the Study

### I.2.i. Study Aims & Expected Outcomes

The aim of this study is to design and implement screening techniques for mutations in the *PKD1* gene. Mutations identified will be assessed in order to determine if significant genotype/phenotype correlations exist in ADPKD. To date, no common pathogenic mutations have been detected in *PKD1* (Krawczak & Cooper, 1997). Since private mutations in the *PKD1* gene are usual, and the size of the population in the present study is relatively small, it is not anticipated that a definitive answer to the genotype/phenotype question will be achieved. However, novel sequence variations discovered here in both the unique and duplicated regions of the *PKD1* gene will add to the growing numbers of *PKD1* mutations and polymorphisms. As greater numbers of mutations are discovered, it will be possible to assess if there is a pre-disposition to particular types of mutation in the gene and if clustering of mutations exist. If mutational clustering does exist, it will be of interest to assess in which functional domains of polycystin-1 it occurs. In addition to improved data on the *PKD1* gene, the detection of familial mutations will also allow diagnostic laboratories to offer molecular diagnosis to pre-symptomatic family members.

### III.2.ii. Study Plan

The investigation of mutations in the *PKD1* gene in this study will be divided into three parts. The first will involve screening of the unique region of the gene by single strand confirmation polymorphism analysis (SSCP) in 48 *PKD1*-linked unrelated individuals from Australia, Bulgaria and Poland. When sequence variations are indicated by SSCP, manual

DNA sequencing will be performed. Manual sequencing was chosen because of its availability. It also has the advantage that heterozygotes are easily detected without the ambiguity often associated with other methods. This will be followed by an assessment of mutation detection efficiency using SSCP compared to direct automated sequencing analysis. The second part of the study will address one of the major problems associated with mutation detection screening of the *PKD1* gene. It will involve the determination of appropriate methods to investigate the duplicated region of the gene without the co-amplification of *PKD1* homologous genes. It will also include investigation of the *PKD1* HGs, to provide information on their functional significance. The final part of the study will be the implementation of these mutation-screening techniques in the duplicated region of the *PKD1* gene. This screening will be performed on the DNA of *PKD1*-linked ADPKD individuals from Australia.

## **II. Review of Literature**

## **II. Literature Review**

### **II.1. General Literature**

#### **II.1.i. Clinical Course and Heterogeneity of ADPKD**

The first major study on ADPKD was based on 284 Danish patients and their families (Dalgaard, 1957). Apart from reporting the prevalence of ADPKD, this study and others since (Churchill *et al.*, 1984; Gabow *et al.*, 1992; Ravine *et al.*, 1992; Coto *et al.*, 1995) have shown marked phenotypic heterogeneity. Differences have also been noted in the prevalence of ADPKD (Yersin *et al.*, 1997; Kirubakaran, 1998). ADPKD appears to be less common in people of African descent than in Caucasians (Yersin *et al.*, 1997). Approximately 65% of the population of the Seychelles are of African descent and 30% are of Caucasian or mixed descent. Of 42 families with ADPKD studied in the Seychelles over a three-year period, only three families were of African descent (Yersin *et al.*, 1997). A low incidence (2/423 renal patients) of ADPKD has also been noted amongst Central Australian Aborigines (Kirubakaran, 1998).

As stated, there is wide variation in age at onset of end-stage renal failure (ESRF) (Milutinovic *et al.*, 1984; Parfrey *et al.*, 1990; Gabow, 1993; Grünfeld, 1998). Most affected individuals reach ESRF between the ages of 45-70 years, and approximately 30-50% never progress to ESRF (Parfrey *et al.*, 1990). Mutations in the *PKD1* gene generally lead to an earlier age at onset and a faster rate of progression towards ESRF than mutations at the *PKD2* locus (Ravine *et al.*, 1992; Hateboer *et al.*, 1999b). A small proportion of affected individuals who show linkage to *PKD1* exhibit very early onset



(Zerres, *et al.*, 1993). Such patients can present *in utero* and, if they survive the first 12 months, ESRF can occur as early as 3 years of age (MacDermot *et al.*, 1998).

Hypertension is the most common complication associated with ADPKD, occurring in approximately 30% of children and 60% of adults prior to the onset of chronic renal failure (CRF) (Gabow, 1993). Hypertension in ADPKD appears to result in an earlier age at onset and an increased rate of progression of renal disease (Milutinovic, *et al.*, 1984; Chapman & Schrier, 1991; Gabow *et al.*, 1992). A multi-centre study (Geberth *et al.*, 1995b) has shown that ADPKD families with a concomitant genetic predisposition to essential hypertension show earlier onset of terminal renal failure. Left ventricular hypertrophy has been shown to occur at a relatively early age in ADPKD hypertensive individuals by comparison with non-ADPKD patients with essential hypertension (Chapman & Schrier, 1991).

Extra-renal manifestations of ADPKD include hepatic, ovarian, seminal vesicular, pancreatic, pulmonary, oesophageal and splenic cysts (Danaci, *et al.*, 1998). Hepatic cysts are the most common and, although not usually present in ADPKD patients at an early age, they are detected in approximately 75-80% of affected individuals aged 60-70 years (Dalgaard, 1957; Milutinovic, *et al.*, 1980; Grünfeld *et al.*, 1985). Cysts in other organs appear to be less common, with little information on their incidence (Gabow, 1993; Danaci, *et al.*, 1998).

Intracranial aneurysms (ICAs) are another well-documented symptom associated with ADPKD, with an estimated incidence of 4-11% (Kaehny & Everson, 1991; Lozano &

Leblanc, 1992; Chapmen *et al.*, 1992; Schievink, *et al.*, 1992; van Dijk *et al.*, 1995; Fick *et al.*, 1995). These figures may be under-estimates, since a prospective Japanese study on 41 ADPKD patients identified ICAs in 17 patients *i.e.* 41% (Wakabayashi *et al.*, 1983). ADPKD has been associated with a higher incidence of several other diseases compared to the incidence found in the general population. These include heart disease, particularly mitral and tricuspid valve prolapse (Hossack *et al.*, 1988), and colonic diverticuli (Starnes *et al.*, 1985; Gabow 1990). ADPKD patients show increased rates of colonic perforation compared to other renal patients (Scheff *et al.*, 1980).

ADPKD co-occurs most commonly with tuberous sclerosis (TSC), an autosomal dominant disorder that results in benign tumours (hamartoma), usually in the brain, skin and kidney. The *TSC2* gene, mutations in which account for approximately 50% of cases of tuberous sclerosis, is located adjacent to *PKD1* in a tail-to-tail orientation on chromosome 16 (Harris, 1997). A gross deletion involving both *TSC2* and *PKD1* was instrumental in mapping the *PKD1* gene (Wunderle *et al.*, 1994), (see section II.2.i).

Other conditions reported to co-occur less commonly with ADPKD are summarised in Table II.1. No genetic cause of co-occurrence has been established between ADPKD and any of these conditions.

The co-existence of cystic fibrosis (CF) and ADPKD may prove to be one of the most interesting associations. Cystic fibrosis, the most common autosomal recessive disease in Caucasians, has a reported incidence in Australia of 1:3,000 births and is characterised by intestinal and pulmonary obstruction (Massie *et al.*, 2000). The

presence of both disorders seems to have an effect on the phenotype of ADPKD (O'Sullivan *et al.*, 1998). In a large family, adult members affected by both ADPKD and CF (but not the CF carriers) displayed less severe polycystic kidney and liver disease (O'Sullivan, *et al.*, 1998).

*In vitro* experiments on monolayer cell cultures from ADPKD renal cystic epithelia, suggest that the CF protein, cystic fibrosis transmembrane conductance regulator (CFTR), is responsible for active chloride ion ( $\text{Cl}^-$ ) and fluid secretion in the cysts (Brill *et al.*, 1996; Davidow *et al.*, 1996; Hanaoka *et al.*, 1996). In pulmonary cells, the CFTR controls the net movement of  $\text{Cl}^-$  across the apical membrane (Bear *et al.*, 1991) and may also be involved in modulating vesicle transportation, organelle function and *Pseudomonas* binding (Bradbury 1999). Mutations in the CFTR gene lead to a decrease in fluid secretion and an increase in fluid absorption (Grantham *et al.*, 1995). It has been proposed that in contrast to CF, where too little  $\text{Cl}^-$  and fluid are secreted, in ADPKD patients too much  $\text{Cl}^-$  and fluid are secreted by cyst epithelia (O'Sullivan *et al.*, 1998). This finding has been supported by other studies (Devuyst *et al.*, 1999; Hanaoka & Guggino, 2000), although the change in expected phenotype may be dependent on the specific CF mutations (Persu *et al.*, 2000). A decline in ADPKD phenotype severity, associated with co-occurring ADPKD and CF, suggests that the modulation of epithelial  $\text{Cl}^-$  secretion may have therapeutic benefit (Cotton & Avner, 1998), but this hypothesis has yet to be tested *in vivo*.

**Table II.1.** Diseases Reported to Co-occur with ADPKD

Disorder	Description	Number of families	References
Caroli's disease	Cystic dilations of the intrahepatic biliary ducts	4	Jordan <i>et al.</i> (1989); Mousson <i>et al.</i> (1997); Torra <i>et al.</i> (1997)
Adenoma of the ampulla of Vater	Causes biliary obstruction	2	Norton <i>et al.</i> (1995); Serafini & Carey (1999)
General overlap connective tissue disorder	Aortic root dilation & aneurysm, aortic valve incompetence, pectus abnormalities, joint laxity, arachnodactyly, scoliosis, dolichostenomella, and high arched palate	4	Somlo <i>et al.</i> (1993); Kaplan <i>et al.</i> (1997)
Marfan syndrome	A connective tissue disorder affecting the skeletal and cardiovascular systems and the eyes	At least 5	DiMatteo <i>et al.</i> , (1965); Selgas <i>et al.</i> , (1981); Ortino <i>et al.</i> , (1988); Biermann <i>et al.</i> , (1992); Hateboer <i>et al.</i> , (2000)
Oral-facial-digital syndrome (type I)	Oral and digital anomalies with facial dysmorphism	Numerous	Harrod <i>et al.</i> (1976); Donnai <i>et al.</i> (1987); Connacher <i>et al.</i> (1987); Feather <i>et al.</i> (1997)
Skeletal malformations	One case with polydactyly and PKD	1	Turco <i>et al.</i> (1993)

### II.1.ii. The Pathogenesis of ADPKD

Altered growth in the kidney epithelial cells of ADPKD patients has been observed under electron microscopy, with cyst wall epithelial cell hyperplasia and micropolypl formation within the cyst observed (Evan *et al.*, 1979). The formation of cysts is not simply caused by an increase in the size of the renal tubules due to the stretching of cells, but results from cell proliferation (Grantham *et al.*, 1987). It has been suggested that, by precipitating renal ischaemia and an increase in renin release, cyst expansion can contribute to the early development of hypertension in ADPKD and thus further accelerate progression to ESRF (Chapman *et al* 1990).

Apoptotic DNA fragmentation has been detected in the polycystic kidneys of 16 patients, 10 of whom had ADPKD and renal failure, 5 had ADPKD but no renal failure, and one was from a one day-old infant with a terminal recessive form of PKD (Woo, 1995). Apoptosis was not observed in the kidneys of 12 normal controls or in the kidneys of 25 renal patients who did not have polycystic disease (Woo, 1995). The results suggest that apoptotic loss of renal tissue may contribute to the progressive deterioration of renal function in PKD (Woo, 1995). Since the gene products of *PKD1* and *PKD2* are believed to be involved in cell-matrix and cell-cell interactions, their deficiency or altered function may play a role in the development of apoptosis seen in ADPKD kidneys (Zhou & Kukes, 1998).

There is evidence that fewer than 1% of renal tubules become cystic and lead to loss of renal function (Grantham, 1997). As cyst numbers increase, kidney function concomitantly deteriorates and serum creatinine levels rise, ultimately leading to ESRF

in 50% of affected individuals by 59 years of age (Parfrey *et al.*, 1990). Kidney cysts have the following characteristic features: (a) they all originate from renal tubular epithelium; (b) as cysts expand, there is a differential increase in the number of epithelial cells found in kidney tubules compared to the number of cells in normal tubules; (c) fluid derived from the glomerular filtrate and net trans-epithelial secretion of solute accumulates in the cyst cavity; and (d) cysts become enlarged in association with matrix re-modelling (Grantham, 1997). Since there is substantial intra- and inter-familial variance in the rate of kidney enlargement, it has been postulated that epigenetic factors play a role in the modulation of renal cyst growth (Grantham, 1997).

Evidence of a neutral lipid, which has the capacity to advance the progression of ADPKD, has been reported (Grantham, 1997). This lipid, or family of lipids, acts as a secretory agonist (secretagogue) and has been detected in the cyst fluids removed from patients with advanced ADPKD and in the cyst fluid of mice with PKD (*pcy/pcy*) (Grantham, 1997). These cyst fluids have also been shown to stimulate the proliferation of MDCK (Madin-Darby canine kidney) renal epithelial cells in culture (Yamaguchi *et al.*, 1995). The secretagogue, referred to as cyst-activating factor (CAF), may stimulate progression of the disease in two ways; (a) by increasing cellular proliferation and fluid secretion, and (b) by increasing interstitial inflammation and fibrosis (Grantham, 1997).

The glycoprotein laminin present in human serum and urine, is another factor that may play a role in cyst proliferation (Slade *et al.*, 1998). Laminin is one of the major intrinsic macromolecules of the tubular basement membrane. In a study by Slade *et al.* (1998), undiluted cyst fluid contained no intact laminin but laminin fragments were present. Fragments of this type, isolated from human placenta, have previously been

shown to increase proliferation of cultured epithelial cells, both normal and ADPKD (Slade *et al.*, 1996). It is believed that contact between laminin fragments and the epithelial cells lining renal tubular cysts could contribute to increased epithelial proliferation in ADPKD (Slade *et al.*, 1998).

The effect of transforming growth factor-alpha (TGF- $\alpha$ ) and epidermal growth factor (EGF), have also been investigated in terms of their effect on ADPKD cell proliferation (Lee *et al.*, 1998). Both of these agents are known mitogens expressed in the kidney, and bind to the same epidermal growth factor receptor (EGFR) (Lee *et al.*, 1998). A study employing immuno-staining showed that renal cystic epithelia exhibit increased expression of both the growth factor TGF- $\alpha$  and its receptor (EGFR) mRNA (Lee *et al.*, 1998). Intense cytoplasmic staining of the tubular epithelia for TGF- $\alpha$  and EGFR proteins has been shown at all stages of cyst development in ADPKD kidney tissues compared to normal human controls (Lee *et al.*, 1998). These elevated levels of TGF- $\alpha$  and EGFR expression were not detected in the surrounding tissues, suggesting that TGF- $\alpha$  and EGFR signalling may be a common mechanism in cystic epithelial proliferation (Lee *et al.*, 1998).

Cysts require a net influx of fluid to prevent collapse of the cell walls (Murcia *et al.*, 1999). Therefore the transition of normal absorptive epithelium into cystic secretory epithelium is necessary for cyst pathophysiology (Murcia *et al.*, 1999). This transition is not irreversible, as cystic epithelia have been shown to revert to wild-type absorptive behaviour in the absence of the stimulating endogenous secretagogues found in cyst fluid, such as forskolin, an adenylate cyclase activator (Grantham *et al.*, 1995).

Monolayer human ADPKD patient-derived cell cultures embedded in type-I collagen have been shown to form cysts in the presence of forskolin or epidermal growth factor (Woo *et al.*, 1994). This cyst formation does not occur when normal kidney cells are treated in the same manner, indicating that mutant ADPKD cells have an intrinsic ability which differs from that of normal kidney epithelial cells (Woo *et al.*, 1994). Cyst formation has been reversibly inhibited *in vitro* by ouabain and amiloride (both are diuretics used as anti-hypertensive drugs), plus colchicine, vinblastine and taxol (these compounds interrupt the cell cycle during mitosis and have been used as anti-cancer drugs) (Woo *et al.*, 1994). Experiments using colchicine, vinblastine and taxol have been conducted *in vivo* on cpk mice, a murine model of autosomal recessive PKD (Woo *et al.*, 1994). Both colchicine and vinblastine proved lethal in the mice, however taxol administered on a weekly basis decreased cyst enlargement, resulted in minimal loss of renal function, and prolonged life from 25-28 postnatal days in untreated mice up to 200 days in taxol-treated mice (Woo *et al.*, 1994). These findings imply that, when the key processes of renal cyst formation and proliferation are understood, dietary and pharmacological treatments may help to delay the proliferation of cysts and the onset of ESRF in ADPKD (Grantham, 1997).



## II.2. The Genetic Basis of ADPKD

### II.2.i. Locus Heterogeneity in ADPKD

The first major breakthrough in the localisation of the ADPKD genes occurred when close linkage to the  $\alpha$ -globin locus on the short arm of chromosome 16 was observed (Reeders *et al.*, 1985). This was achieved using a highly polymorphic DNA marker (the 3' HVR-hypervariable region) approximately 8 kb beyond the 3' end of the  $\alpha$ -globin gene. The marker was analysed in four British and five Dutch families totalling 183 members, and gave a maximum combined lod score of 25.85 at  $\theta = 0.05$  (99% confidence interval 2-11 cM) (Reeders *et al.*, 1985). Each of the nine families displayed linkage to this locus, now known to be the *PKD1* gene. The families presented with the most typical clinical form of the disease, ESRF occurred at a mean age of 48 years and, in the effective absence of phenotypic variation, genetic heterogeneity was deemed improbable (Reeders *et al.*, 1985). A subsequent study in South Wales of 12 ADPKD kindreds gave a maximum combined lod score of 24.187 ( $\theta = 0.03$ ) using the  $\alpha$ -globin and 3' HVR probes (Lazarou *et al.*, 1987).

Refined linkage mapping became possible shortly afterwards with the marker 24-1, which flanks the *PKD1* gene on the side opposite to the  $\alpha$ -globin gene (Breuning *et al.*, 1987). A joint study analysing 27 Northern European ADPKD families failed to detect evidence of linkage heterogeneity (Reeders *et al.*, 1987). A note by Kimberling and Gabow added to this paper in proof, suggested possible locus heterogeneity in families with more diverse ethnic origins. The first example was an ADPKD family of Sicilian origin which did not show linkage to the  $\alpha$ -globin gene on chromosome 16p

(Kimberling *et al.*, 1988). An average of 85% of all Caucasian families with ADPKD show linkage to *PKD1* (Peters & Sandkuijl, 1992). For that reason this is the major gene selected for investigation in this thesis.

The position and order of the many markers isolated around the *PKD1* region was initially determined using a panel of somatic cell hybrids and extensive linkage analysis in normal and *PKD1* families (Breuning *et al.*, 1990; Reeders *et al.*, 1988; Germino *et al.*, 1990). Further work creating a long-range restriction map localised the *PKD1* region to an area spanning approximately 750-500 kb (Harris *et al.*, 1990; Germino *et al.*, 1993), flanked by the markers GGG1 and SM7 (Harris *et al.*, 1991; Somlo *et al.*, 1992). The region has a high CpG content and could contain as many as 20 genes. This large number of candidate genes hindered the subsequent fine localisation of the *PKD1* gene (Germino *et al.*, 1993).

Linkage disequilibrium studies were undertaken using additional markers: VK5B in a Scottish study (Pound *et al.*, 1992), and BLu 24 with a Spanish population (Peral *et al.*, 1994). These studies showed evidence of a common ancestor in a proportion (approx. 40%) of each population, but the degree of linkage disequilibrium was too weak to position the *PKD1* locus with precision.

As previously noted, the first ADPKD family which did not display linkage to the alpha-haemoglobin complex on chromosome 16p was of Sicilian origin (Kimberling *et al.*, 1988), and was clinically indistinguishable from families in which ADPKD was linked to chromosome 16 (Kimberling *et al.*, 1988). Numerous families were

subsequently reported in the literature to have a form of ADPKD that did not co-segregate with genetic markers on chromosome 16p (Elles *et al.*, 1990; Bachner *et al.*, 1990; Coto *et al.*, 1992; Bear *et al.*, 1992). As more non-*PKD1*-linked families were identified it was noted that ADPKD in these families appeared to exhibit a milder clinical phenotype (Bear *et al.*, 1992; Ravine *et al.*, 1992).

The second ADPKD gene, *PKD2*, was mapped in the original Sicilian kindred to a 9-cM segment flanked by the markers D4S231 and D4S414 on the long arm of chromosome 4 (Kimberling *et al.*, 1993). The highest lod scores in this two-point linkage analysis were obtained between ADPKD and D4S395 ( $Z = 9.83$ ,  $\theta = 0.07$ ) and ADPKD and D4S414 ( $Z = 10.13$ ,  $\theta = 0.06$ ). Peters *et al.* (1993) independently mapped the *PKD2* gene to chromosome 4q, in a study of eight families where ADPKD did not co-segregate with markers on chromosome 16p. This chromosome had previously been studied for Huntington disease (Gusella *et al.*, 1983) and facioscapulohumeral dystrophy (Wijmenga *et al.*, 1991), so markers closely linked to *PKD2* were readily available for predictive diagnostic purposes (Peters *et al.*, 1993).

The question of further locus heterogeneity in ADPKD remains open, since at least five families show linkage to neither *PKD1* nor *PKD2* (Daoust *et al.*, 1995; Bogdanova *et al.*, 1995; de Almeida *et al.*, 1995; Turco *et al.*, 1997; Ariza *et al.*, 1997). A possible positive lod-score to markers linked to chromosome 7 has been reported for "PKD3" (de Almeida *et al.*, 1999). Phenotypic variation has been found between the five families, which may result from modifying or environmental factors, as suggested in *PKD1* and *PKD2* (Hateboer, *et al.*, 1999a). Alternatively, there may be more than three

genes responsible for the ADPKD phenotype, although the existence of even a third locus for ADPKD has been questioned since many confounders can lead to false exclusion of linkage (Paterson & Pei, 1998).

It was suggested that bilineal transmission of two independent PKD mutations within the same family might lead to false exclusion of linkage when only one disease locus is considered at a time (Paterson & Pei, 1998). This could arise either through mutations involving both copies of *PKD1* or *PKD2* segregating within the same family. In such cases, individuals carrying two identical mutations are referred to as homozygotes, whereas those with two different mutant alleles are termed compound heterozygotes. Alternately, one *PKD1* and one *PKD2* mutation could segregate within the same family. In such cases, some affected individuals would have digenic mutations (*i.e.* in both ADPKD genes) and others would have only *PKD1* or *PKD2* mutations. Predictably these events would be rare. However, only five ADPKD families have failed to show linkage to either the *PKD1* or *PKD2* genes (Paterson & Pei, 1998). In 2001 the first family with bilineal ADPKD was reported (Pei *et al.*, 2001). Family members had either germline mutations in *PKD1* or *PKD2*, and a smaller, more severely affected subset were shown to have both mutations.

A novel *PKD2*-like gene (*PKDL*) has been reported (Nomura *et al.*, 1998) which displays 50% amino acid (aa) sequence identity and 71% homology to *PKD2*. The structure of the gene is suggestive of a role as a sub-unit of an ion channel, but there is no evidence of linkage to suggest that it could be associated with ADPKD in humans (Nomura *et al.*, 1998). *PKD2L2* is a third gene with significant homology to *PKD2* and

*PKDL*. It has been recently mapped to human chromosome 5q31, and the gene product (polycystin-L-2) has 48% identity with polycystin-2 and 47% identity with polycystin-L (Guo *et al.*, 2000).

#### II.2.ii. Ethnic Heterogeneity of ADPKD

Two Southern European groups have reported a lower incidence of linkage to the *PKD1* gene than had previously been reported. A Bulgarian study of 22 families showed approximately 68% of patients with linkage to *PKD1* (Bogdanova *et al.*, 1995), and a smaller study of seven Greek-Cypriot families showed linkage to *PKD1* in only four families, the others were linked to *PKD2* (Constantinou-Deltas *et al.*, 1995). Due to the small numbers in each study, it is not possible to determine if locus heterogeneity is significantly different when compared to the larger study populations previously conducted in Northern Europeans (Peters & Sandkuijl, 1992). A more recent study on 48 ADPKD families of Korean descent found that 79% of individuals had ADPKD linked to *PKD1*, and 21% were linked to *PKD2* (Lee *et al.*, 2001).

#### II.2.iii. Molecular Cloning of the PKD Genes

Cloning of the *PKD1* gene was facilitated by the identification of a family with two genetic disorders, both due to mutations in genes located on chromosome 16p13.3 (Wunderle *et al.*, 1994). Affected members in a Portuguese kindred had two distinct phenotypes, either typical autosomal dominant polycystic kidney disease (*PKD1*) or type-2 tuberous sclerosis (*TSC2*). The mother and daughter in this family had balanced translocations, 46XX t(16;22)(p13.3;q11.21) and displayed ADPKD but not *TSC2*. In contrast, the son had an unbalanced translocation 45XY/-16-22+der(16)

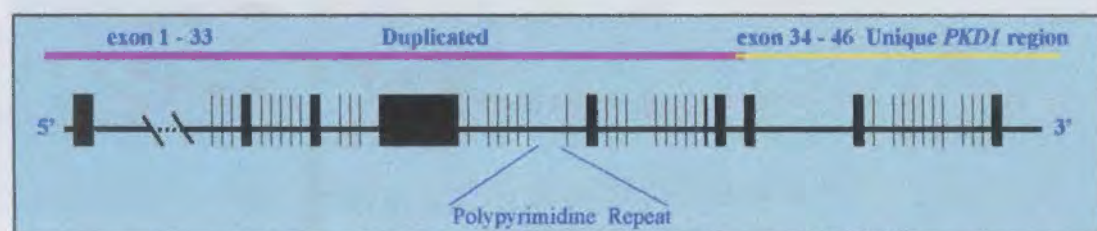
(16qter→16p13.3::22q11.21→ 22qter) and displayed the clinical features of *TSC2* but not ADPKD. It was therefore concluded that the unbalanced translocation most probably resulted in a deletion at the *TSC2* locus, whereas the balanced translocation was investigated to see if it disrupted the *PKD1* locus (The European PKD Consortium, 1994). Investigation of the family using polymorphic markers refined the *PKD1* locus to 60 kb and enabled the construction of a cosmid contig covering the breakpoint region. Hence the polycystic breakpoint (PBP) gene was listed as the most likely candidate gene for *PKD1* (The European PKD Consortium, 1994).

More recently in humans the polycystic kidney disease and receptor for egg jelly related gene (*PKDREJ*) was localised close to the same region of chromosome 22 (Hughes *et al.*, 1999). The PKDREG protein is expressed exclusively in the testis of humans coincident with sperm maturation, suggesting a role in the fertilisation process (Hughes *et al.*, 1999).

The *PKD2* gene was cloned by the construction of a cosmid- and P1-based contig extending over 680 kb on chromosome 4q. Northern blot hybridisation revealed a 5.4 kb transcript (cTM-4) expressed in most fetal and adult tissues, including kidney, ovary, testis, small and large intestine and fetal lung (Mochizuki *et al.*, 1996), which showed homology at the amino acid level with the *PKD1* gene product. Mutation screening in the cTM-4 transcript in three unrelated affected individuals revealed three nonsense mutations that segregated with the disease phenotype in their families, and on that basis cTM-4 was deemed to be the *PKD2* gene (Mochizuki *et al.*, 1996).

## II.2.iv. Structure of the *PKD1* Gene

As previously stated, the *PKD1* gene comprises 46 exons spanning 52 kb, and produces a 14.5 kb mRNA transcript which encodes a 4302 aa protein with novel domain architecture (The International PKD Consortium, 1995). The *PKD1* gene contains 3 long polypyrimidine tracts in introns 1, 21 and 22 (Van Raay *et al.*, 1996), the longest of which (depicted in Fig II.1) is located in IVS 21 and spans 2.5 kb.



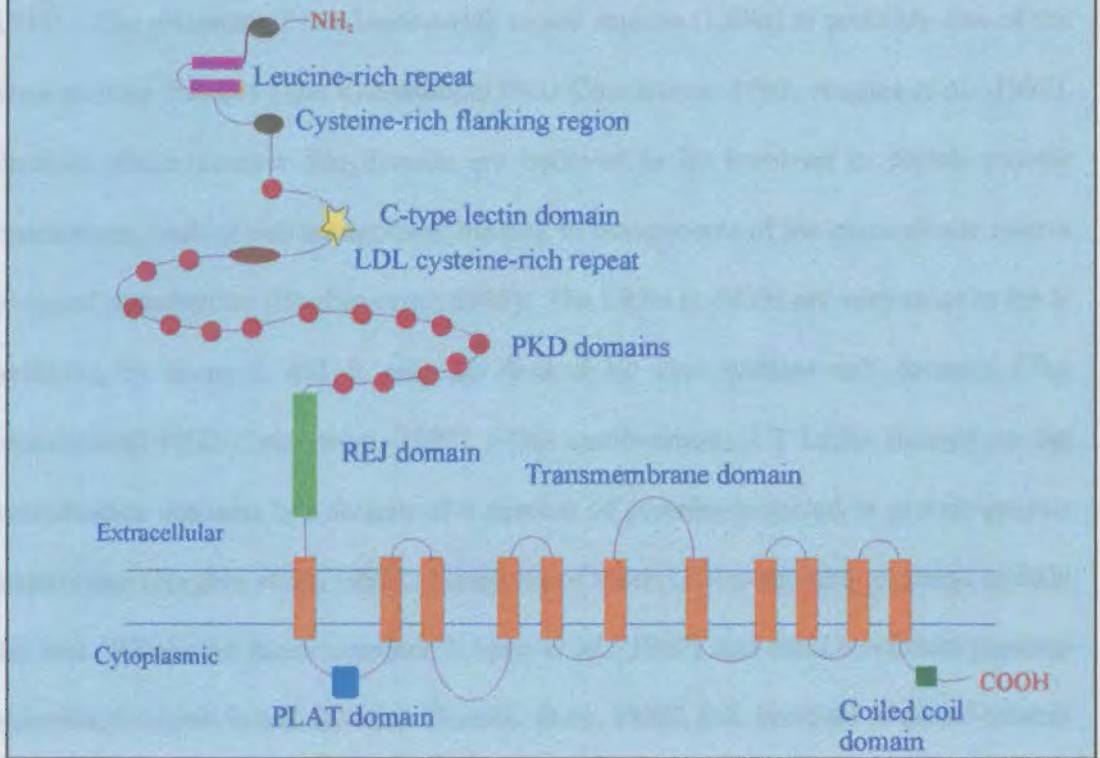
**Figure II.1.** The *PKD1* Gene. (Adapted from Hughes *et al.*, 1995).

## II.2.v. The *PKD1* Protein

Polycystin-1, the predicted *PKD1* protein, is a glycoprotein which consists of an extracellular 5' N-terminus of over 2500 aa, multiple transmembrane domains, and a cytoplasmic 3' C-terminus (Fig II.2) (The International PKD Consortium, 1995; Hughes *et al.*, 1995).



# Polycystin-1



**Figure II.2.** Polycystin-1.

The basic defect in ADPKD is believed to be tubular dysmorphogenesis. The principal features include altered structure and function of the Golgi complex, altered structure and composition of the matrix, and abnormal cell differentiation, resulting in faulty matrix-cell communication (Carone *et al.*, 1995). It has therefore been proposed that polycystin-1 is either involved in the mediation of cell-cell or cell-matrix interactions, or that it could be an intrinsic component of the extracellular matrix (The International PKD Consortium, 1995).



The domain architecture of polycystin-1 has been predicted by several groups, with only slight suggested variations (The International PKD Consortium, 1995; Hughes *et al.*, 1995). The presence of two leucine-rich repeat regions (LRRs) is probably one of the most striking features (The International PKD Consortium, 1995; Hughes *et al.*, 1995). Proteins which contain this domain are believed to be involved in protein-protein interactions, such as cell recognition, binding to components of the extracellular matrix or signal transduction (Hughes *et al.*, 1995). The LRRs in *PKD1* are very close to the 5' terminus, in exons 2 and 3, and are flanked by two cysteine-rich domains (The International PKD Consortium, 1995). This combination of 2 LRRs flanked by the cysteine-rich domains is a feature of a number of proteins involved in protein-protein interactions (Hughes *et al.*, 1995). Examples of these LRR-containing proteins include the von Willebrand factor receptor (Lopez *et al.*, 1987) and the *Drosophila* proteins chaoptin, involved in cell adhesion (Reinke, *et al.*, 1988), toll, involved in dorsal-ventral polarity (Hashimoto *et al.*, 1988), and slit, involved in morphogenesis (Rothberg *et al.*, 1990). Homology has also been observed between the *PKD1* exon 4 encoded domain and the *trk* gene (Hughes *et al.*, 1995), a proto-oncogene known to encode a membrane-spanning tyrosine kinase expressed in neural tissues (Martin-Zanca *et al.*, 1986).

The next functional domain in *PKD1* is a C-type lectin domain. The usual function of the domain, associated with cell adhesion and recognition, is to bind specific carbohydrates in the presence of calcium ions. In most proteins, this domain is encoded by either one or three exons (Drickamer, 1989), *e.g.* the human *LOX-1* gene (Aoyama *et al.*, 1999) and the C-type lectin superfamily cartilage derived gene, *CLECSF1* (Neame

*et al.*, 1999), both of which have three exons for this domain. In *PKDI* the domain is not typical, as it is encoded by a novel number of exons (exons 6 and 7), in association with a previously unreported combination of protein domains (Hughes *et al.*, 1995).

The most numerous of the *PKDI* domains is a novel Ig-like repeat, dubbed the PKD domain (The International Consortium, 1995). According to this source, the Ig-like domain is repeated 14 times, once in exon 5, once in exon 13, once in exon 14 and 11 times in exon 15 (aa 1031-2142). The Ig-like repeat was detected 16 times by Hughes *et al.* (1995), who placed the first repeat in exon 5 and the remaining 15 repeats in exons 11-15. Both groups state that the repeat motif consists of about 80-90 aa and, although the function of the Ig-like repeats has not been established, they too are in an extracellular position. The area covered by these repeats represents approximately 30% of the *PKDI* gene. The current opinion is that there are 16 copies of the repeat (van Adelsberg, 1999). The Ig-like repeats have been shown to have a Greek key  $\beta$ -sandwich topology (Bycroft *et al.*, 1999), also known as the immunoglobulin (Ig)-like fold (Bork *et al.*, 1994). It is built from two  $\beta$ -sheets, one with three strands (labelled A, B, and E) and the other with four strands (labelled G, F, C, and C'), packed face to face. The structure of the Ig-like repeat found in *PKDI* is similar to the Ig-like fold found in other cell surface proteins, however it is not identical to any known protein and therefore represents a novel protein family (Bycroft *et al.*, 1999).

Repeat number ten is believed to be of particular functional significance, as all of its surface residues are conserved on one  $\beta$ -sheet in the *Fugu* genome (Bycroft *et al.*, 1999). Comparative analysis of the human and *Fugu* genomes for *PKDI* has

demonstrated *PKDI* repeat number 10 to be in an evolutionarily conserved synteny group, suggesting that it may play an important role in normal cellular physiology (Sandford *et al.*, 1997). The 40% sequence identity between human and *Fugu* polycystin-1 suggests that it is the tertiary structure of this protein, rather than sequence identity, which is necessary for normal function (Sandford *et al.*, 1997).

The next domain, FN-III, shows extensive homology to the sea urchin receptor for egg jelly (REJ) protein (Moy *et al.*, 1996). REJ, a 210-kD sperm membrane glycoprotein, has been shown to function as an ion channel regulator and is believed to trigger the sea urchin sperm acrosome reaction (Moy *et al.*, 1996). The REJ protein is modular in design and consists of one EGF module (containing the 7 diagnostic Cys residues), two C-type lectin carbohydrate recognition modules and a novel module, 'the REJ module', which shares homology with the human *PKDI* protein. The REJ module comprises 50% of the REJ protein (residues 480-1187) and 17% of polycystin-1 (residues 2146-2882). The two modules (Altschul *et al.*, 1990; Pearson & Lipman, 1988) are 20% identical (Moy *et al.*, 1996). The evolutionary homology between the REJ of sea urchin and human *PKDI* suggests that *PKDI* may also be involved in ion channel regulation (Vacquier & Moy, 1997).

There is some inconsistency in the literature concerning the other domains of the *PKDI* gene. The International PKD Consortium (1995) reported a low-density lipoprotein type-A (LDL-A) domain in exon 10. Due to its hydrophobic nature, the domain was predicted to have a receptor-related role as a ligand-binding region. Hughes *et al.* (1995) failed to mention such a domain but they did predict the presence of four

fibronectin type III (FN-III) repeats, situated beyond the Ig-like repeats toward the 3' C-terminal region of polycystin-1. These domains are found in extracellular matrix proteins, and in extracellular locations of receptor and cell adhesion molecules (Hughes *et al.*, 1995).

Each of the groups studying polycystin-1 also indicated the presence of several (most probably 11) transmembrane domains (Hughes *et al.*, 1995; The International PKD Consortium, 1995). The region around the transmembrane (TM) domains (numbers TM1-6 and TM11) has recently been subjected to further analysis (Bateman & Sandford, 1999). This investigation was undertaken using the PSI-BLAST sequence comparison program, on the premise that there is a high degree of sequence conservation between human and *Fugu* polycystin-1. This sequence region was found to have matches to 67 other sequences in SWISS-PROT version 37 and TrEMBL version 9 (Bairoch & Apweiler, 1999), namely mammalian lipoxygenases, triacylglycerol lipase and lipoprotein lipase. Following this work, a new protein domain, the PLAT domain (after polycystin-1, lipoxygenase and alpha toxin) has been described (Bateman & Sandford, 1999). The PLAT domain is found in the first intracellular region of the *PKD1* gene between TM1-TM2, corresponding to exons 26-29 (Bateman & Sandford, 1999). The domain is associated with protein-protein and protein-lipid interactions in other proteins, suggesting that the first cytoplasmic loop of polycystin-1 may be important in mediating interactions with other membrane protein(s) involved in polycystin-1 function (Bateman & Sandford, 1999). All of the polycystin-1 domains predicted to date indicate a role in protein:protein signalling (Harris, 1999).

#### II.2.vi. *PKDI* Homologous Transcripts

In addition to the *PKDI* gene, three other homologous genes (HGs) have been identified by Northern blotting (The International PKD Consortium, 1995; Hughes *et al.*, 1995; The European PKD Consortium, 1994). These genes, named HG-A, HG-B, and HG-C, produce transcripts of approximately 21 kb, 17 kb and 8.5 kb respectively (Hughes *et al.*, 1995). As noted previously, the HGs are up to 97% homologous with the *PKDI* gene over two-thirds of its length and it has not been determined if the HG transcripts produce functional proteins (Hughes *et al.*, 1995). The origins of these HGs could be explained in two ways: either they are alternatively spliced forms of the PBP (*PKDI*) gene or they are separate genes co-located in 16p13.1. All three of the identified HGs map to chromosome 16p13.1 proximal to the *PKDI* region (The European Polycystic Kidney Disease Consortium, 1994).

#### II.2.vii. Polycystic Kidney Disease in Animals

The ADPKD murine gene (*Pkd1*) has been mapped to mouse chromosome 17 (Olsson *et al.*, 1996). Like its human counterpart, it lies tail to tail with the murine *Tsc2* but is not duplicated elsewhere on the murine genome, and does not include the polypyrimidine tracts in introns 21 and 22, suggesting that these changes are recent evolutionary events (Piontek & Germino, 1999). The expression of *Pkd1* has been observed from the morula embryonic stage up to adulthood (Guillaume *et al.*, 1999). High levels of *pkd1* protein expression in the embryo suggest a role in development (Guillaume *et al.*, 1999). Phenotypic variation had been observed between two reported *Pkd1* murine models (Lu *et al.*, 1997; Kim, *et al.*, 2000a). The first *Pkd1* model (*PKDI*  $\Delta$ 34) involves a frameshift mutation (Lu *et al.*, 1997), which produces a less severe

phenotype than that observed in the L3946\* mutant polycystin 1 model, where the mouse polycystin-1 is expected to terminate after amino acid L3946 (equivalent to polycystin-1 L3955 in humans) (Kim, *et al.*, 2000a). Other animal models which may be useful for the study of ADPKD are the Persian cat (Feldhahn, 1995; Biller *et al.*, 1996) and the bull terrier dog (O'Leary *et al.*, 1999).

#### II.2.viii. The Structure of the *PKD2* Gene

*PKD2* has been shown to consist of at least 15 exons (Hayashi *et al.*, 1997), and its architecture includes six membrane-spanning domains and intracellular amino- and carboxy- termini (Mochizuki *et al.*, 1996). There is approximately 25% identity between the putative 968 aa *PKD2* gene product, polycystin-2, and almost 450 aa of polycystin-1 (Mochizuki *et al.*, 1996). This homology corresponds to transmembrane domains 1-6 of *PKD2* and covers four transmembrane domains of *PKD1*. The *PKD2* gene is not duplicated on chromosome 4q as *PKD1* is on chromosome 16. The domains that are found in *PKD2* share homology with a family of  $\text{Ca}^{2+}$  (and  $\text{Na}^{2+}$ ) channels, mostly with the L-type voltage-activated  $\text{Ca}^{2+}$  channel  $\alpha_{1E}$  (VACC  $\alpha_{1E}$ ) (Mochizuki *et al.*, 1996).

The murine *Pkd2* gene has been reported by two independent groups (Wu *et al.*, 1997; Pennekamp *et al.*, 1998). *Pkd2* is located on mouse chromosome 5 spanning at least 35 kb and, like its human homologue, consists of 15 exons (Pennekamp *et al.*, 1998). The peptide shows 95% homology to human polycystin-2, with well-conserved functional domains (Pennekamp *et al.*, 1998).

The most interesting aspect of the *PKD2* gene structure relates to its apparent interaction with *PKD1*. A coiled-coil domain in the C-terminus of the *PKD1* gene has been described which binds specifically to the C-terminus of the *PKD2* gene product (Qian *et al.*, 1997). This heterodimeric interaction suggests that *PKD1* and *PKD2* may interact *in vivo* in a signalling pathway for normal tubular morphogenesis. Mutations in either gene could then lead to the identical pathology observed in ADPKD (Qian *et al.*, 1997). The interaction of *PKD1* and *PKD2* results in the up-regulation of *PKD1* expression, suggesting that *PKD1* may require *PKD2* for the maintenance of normal expression levels (Tsiokas *et al.*, 1997). Inactivation of the heterodimeric-signalling pathway may impair terminal differentiation and lead to cysts (Arnould *et al.*, 1999). Given evidence of the five known polycystin-type proteins, it is possible that more than two polycystin proteins interact in such signalling pathways (Nomura *et al.*, 1998).

#### 11.2.ix. Expression Patterns of Polycystin-1 and Polycystin-2

Polycystin-1 is expressed in a number of tissues, including the thymus, thyroid, adrenal glands, tonsils, duodenum, stomach, liver, spleen, heart, lungs, testes, brain and kidney (Ward *et al.*, 1996), and it is present from fetal to adult life (Peters *et al.*, 1996). Polycystin-1 has been detected at higher concentrations in cystic kidneys than in normal controls (Ward *et al.*, 1996). In both normal and cystic kidneys, polycystin-1 expression is localised to the epithelium of all tubular structures and the glomerular parietal and visceral epithelium (podocytes) (Peters *et al.*, 1996). Fetal expression of polycystin-1 is higher than the normal adult level (Griffin *et al.*, 1996; Wilson, 1997; Huan & van Adelsberg, 1999; Weston *et al.*, 1997). Some cultured kidney cyst-lining epithelial cells have shown strong staining for polycystin-1 at the site of inter-cell

interaction, adding evidence to the hypothesis that polycystin-1 is involved in cell:cell interactions (Veldhuisen *et al.*, 1999). More recent studies by the same group have shown that polycystin-1 expression is not confined to epithelial cells but is also present in endocrine cells involved in the secretion or transport of molecules. Much stronger staining for polycystin-1 was detected in cardiac muscle compared to skeletal muscle. Cardiac muscle fibres communicate via gap junctions, and the high expression of polycystin-1 may correlate with the number of interactions between these cells (Peters *et al.*, 1999).

With the exception of Ibraghimov-Beskrovnaya *et al.* (1997), all studies (Ward *et al.*, 1996; Griffin *et al.*, 1996; Palsson *et al.*, 1996; Peters *et al.*, 1996; Geng *et al.*, 1996; Wilson, 1997; Huan & van Adelsberg, 1999; Weston *et al.*, 1997; van Adelsberg *et al.*, 1997), have reported that polycystin-1 is expressed at higher levels in the cystic kidney epithelia of adult ADPKD individuals than in normal adult kidneys (van Adelsberg *et al.*, 1999). This contrasts with what one would expect due to the haploinsufficiency associated with an autosomal dominant disorder, or indeed with the two-hit model of cystogenesis (see section II.3.i.). A higher level of expression in affected individuals may be the result of an accumulation of non-functional protein caused by truncating or missense mutations in the *PKD1* gene. Such increased levels of protein expression have been associated with mutations in the *NEM1* and *TPM3* genes linked to nemaline myopathy (Tan *et al.*, 1999).

Polycystin-2 presents with a similar pattern of expression to polycystin-1 in the distal tubule and collecting duct kidney epithelia, but with more marked vascular endothelial



cell staining (Foggensteiner *et al.*, 2000). Co-localisation of the two polycystins was demonstrated by dual colour immunofluorescence (Ong *et al.*, 1999). A 34 aa region of polycystin-2 (Glu<sup>787</sup>-Ser<sup>820</sup>) had been localised to the endoplasmic reticulum of endothelial cells (HEK293) in culture, where it was the site of specific interaction with an as yet unidentified protein (Cai *et al.*, 1999). The localisation of full-length polycystin-2 to intracellular membranes suggested that this protein may be a sub-unit of intracellular channel complexes (Cai *et al.*, 1999).

## II.3. The Mutations and Polymorphisms Associated with ADPKD

### II.3.i. Mutational Mechanisms of ADPKD

In ADPKD, it is important not only to understand how cystic proliferation occurs, but also to appreciate the underlying genetic mechanisms and the factors causing their initiation. A two-hit model has been proposed to explain cyst formation (Qian *et al.*, 1996), involving the hereditary germline mutation (first hit) and a somatic mutation of the normal allele (second hit) which results in the loss of heterozygosity (LOH) of the normal PKD gene product (Qian *et al.*, 1996). The two-hit model was first proposed to explain the genetic origin of retinoblastoma (Knudson, 1971), a form of cancer affecting the light-sensitive retinal cells. Unilateral and segmentally localised forms of PKD give credence to the two-hit hypothesis of ADPKD (Gouldesbrough & Fleming, 1998). Individuals with the disorder have segmental cystic abnormalities in one kidney that are morphologically identical to ADPKD, however they have no family history of the disease and simple surgical resection is curative (Gouldesbrough & Fleming, 1998). If the two-hit model explains cyst formation in ADPKD, it could be postulated that two somatic mutations could lead to the rare disease phenotype in unilateral and segmentally localised PKD (Gouldesbrough & Fleming, 1998).

A number of additional studies have shown that LOH is associated with ADPKD in individual renal cysts (Brasier & Henske, 1997; Koptides *et al.*, 1998; Torra *et al.*, 1999). Three different somatic mutations, along with a missense germline mutation in exon 31 of *PKD1*, were detected in 3 of 17 renal cysts removed from an ADPKD patient during transplantation surgery (Koptides *et al.*, 1998). As previously stated,

accumulation of non-functional protein may occur in these cells and could account for the over-expression of polycystin-1 often associated with ADPKD. Renal cysts associated with a germline mutation in the *PKD2* gene (Koptides *et al.*, 1999; Torra *et al.*, 1999), and hepatic cysts of ADPKD individuals (Watnick *et al.*, 1998b) have also been shown to involve somatic mutations, extending the two-hit hypothesis of cyst formation to the *PKD2* gene and to other organs apart from the kidney. This homozygous loss of normal *PKD1* or *PKD2* function presumably alters the ability of epithelial cells to retain tissue organisation, resulting in proliferation and cyst formation (Wu & Somlo, 2000). In a murine model (WS25) carrying an unstable allele, which introduced an in-frame stop codon nine codons downstream of the 5' cloning junction, there was a cessation in *Pkd2* protein expression in renal epithelial cells after cyst formation (Wu *et al.*, 1998). A true null allele mouse model (WS183) was also observed, and in this model *Pkd2*<sup>-/-</sup> mice were not viable (Wu *et al.*, 1998).

A paper on polycystin-2 expression confirmed the positive staining for polycystin-2 in individual kidney cysts from *PKD2*-linked individuals, thus questioning the two-hit hypothesis as the pathological factor in cyst formation (Ong *et al.*, 1999). It was argued that if the staining is positive for a mutant protein inactivated by a missense mutation, then the two-hit hypothesis is plausible (Ong *et al.*, 1999). Alternatively, the detection of polycystin-2 in cystic epithelia may indicate that cyst initiation can occur without loss of the normal protein or possibly through somatic mutations at other loci (such as *PKD1*), affecting proteins which may be part of the polycystin complex (Ong *et al.*, 1999).

A dominant negative effect of the mutated allele has been proposed (Roelfsema *et al.*, 1997). This occurs when the mutated protein interacts with the wild-type gene product and adversely affects the normal function of that protein. It is suggested that the model applies to a number of autosomal dominant diseases, including generalised resistance to thyroid hormone (GRTH) (Yen *et al.*, 1992), Marfan syndrome (Aoyama *et al.*, 1993), Diabetes–insulin receptor mutation (Levy-Toledano *et al.*, 1994), and growth hormone deficiency (Hayashi *et al.*, 1999). This form of mutational mechanism has also been shown to be associated with a variable phenotypic effect. Differing mutations in the *c-erbA-beta-1-thyroid-hormone* receptor have divergent effects on dimerization activity, and appear to influence the level of dominant negative activity in humans (Hao *et al.*, 1994).

Gene conversion may be a mutational mechanism in ADPKD. One group proposed that the HGs of *PKD1* might represent reservoirs for mutation generation in ADPKD via gene conversion (Watnick *et al.*, 1998a). Gene conversion was first studied in yeast and has been defined as the non-reciprocal exchange of genetic information (Murti *et al.*, 1994). It has been postulated that gene conversion and recombination may be related processes that involve pairing of homologous sequences, although in gene conversion genetic information is transferred from the donor gene to the recipient without the donor undergoing concomitant modification (Murti *et al.*, 1994).

The role of gene conversion as a mutational mechanism has been linked to other human genes and their pseudogenes, *e.g.*, with congenital adrenal hyperplasia (*CYP21* gene) (Collier *et al.*, 1993), neurofibromatosis (Hulsebos *et al.*, 1996), and the von Willebrand

factor gene (Eikenboom *et al.*, 1994). *CYP21* is somewhat similar to *PKDI*: both genes are duplicated on the same chromosome and share greater than 95% sequence identity across many thousands of base pairs with their respective homologue(s) (Collier *et al.*, 1994; Tusie-Luna & White, 1995). In the *PKDI* gene, gene conversion is proposed as a likely cause of the cluster of missense mutations in exon 23 (Watnick *et al.*, 1998a). It is yet to be determined if this mutational mechanism is responsible for germline mutations in other duplicated parts of the gene, or for somatic mutations. Given the fact that the HGs extend over approximately 70% of *PKDI*, gene conversion could well be a major mechanism of mutation (Watnick *et al.*, 1998a).

Another possible source of mutation in the *PKDI* gene is the polypyrimidine tract in IVS 21 (Van Raay *et al.*, 1996), because of its many triple helices (Watnick *et al.*, 1998a). The tract (~2.5 kb) contains a number of repeat elements that are predicted to form H-DNA structures. These structures form a triple helix and have been shown to promote mutagenesis in cultured cells (Wang *et al.*, 1996a). Triplex formation by a poly(dG)-poly(dC) tract separating two homologous sequences in a bacterial system was shown to promote recombination when transcription of the downstream gene was activated (Wang *et al.*, 1996a; Kohwi & Panchenko, 1993). Since recombination and gene conversion are related processes, it might be expected that the polypyrimidine tracts in *PKDI* increase the rate of mutagenesis (Watnick *et al.*, 1998a). Preliminary evidence supporting this theory comes from the potential clustering of missense mutations in exons 23 and 25 of *PKDI* (Watnick *et al.*, 1997; Peral *et al.*, 1997), around the polypyrimidine tracts in introns 21 and 22.

### 11.3.ii. Mutations Found in *PKD1*

As of February 2002 there have been 185 mutations reported in the gene (Table 11.2).

**Table 11.2.** *PKD1* Mutations by Type and Region

<b>Mutation Type</b>	<b>Total Number</b>	<b>3' Unique Region</b>	<b>5' Repeated Region</b>
<b>Nucleotide Substitutions</b>			
missense	53	15	39
nonsense	45	19	26
splicing	15	6	9
<b>Deletions, Insertions and Duplications</b>			
Small deletions	43	12	32
Small insertions	16	10	7
Large deletions	9	4	5
Large duplication	1	1	0
<b>TOTAL</b>	<b>185</b>	<b>67</b>	<b>118</b>

(Krawczak & Cooper, 1997; Afzal *et al.*, 2000; Aguiari *et al.*, 2000; Kim, *et al.*, 2000b; Koptides *et al.*, 2000; Perrichot *et al.*, 2000 a & b; Phakdeekitcharoen *et al.*, 2000 & 2001; Rossetti *et al.*, 2001; Mizoguchi *et al.*, 2001; Bouba *et al.*, 2001; Tsuchiya *et al.*, 2001; Ding *et al.*, 2002).

### 11.3.iii. Types of Mutations in *PKD1*

Approximately 80% of all mutations in the 3' region and 67% of mutations in the 5' region are truncating. This figure is probably biased as PTT was used by two groups to screen for mutations (Peral *et al.*, 1997; Roelfsema *et al.*, 1997). Mutations have been detected in exons 1 to exon 46, although few groups have screened the entire region (Rossetti *et al.*, 2001; Phakdeekitcharoen *et al.*, 2001), and a bias exists for mutation screening in exons 15, 23 and 25 (Table 11.3). As more studies screen the entire 5' region (exons 1 to 33) for both truncating and non-truncating mutations, a true indication of the specific types of mutation involved will be established for the *PKD1* gene.

### 11.3.iv. Location of *PKDI* Mutations

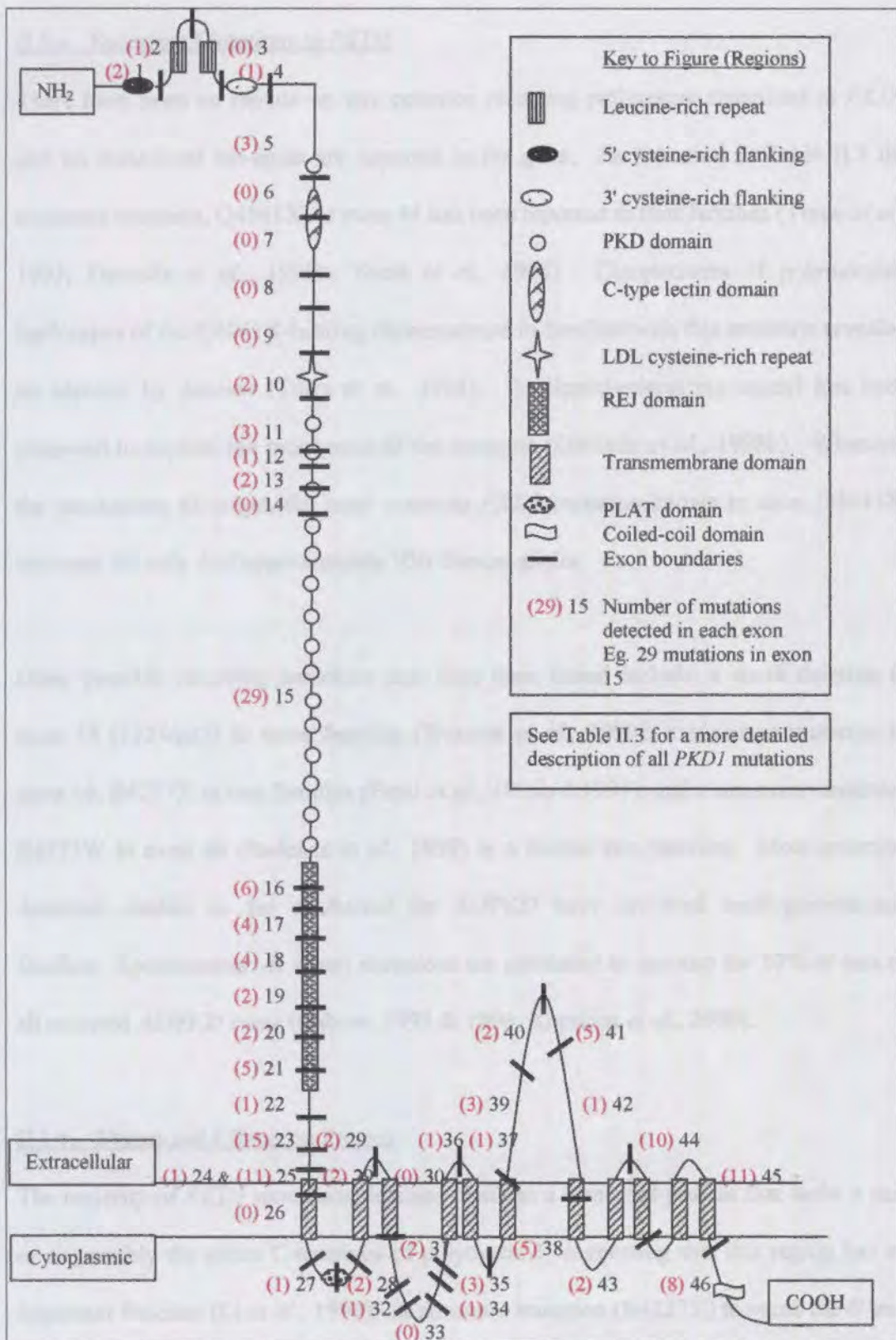
A comprehensive list of mutations published in the *PKDI* gene are given in Table 11.3, and their locations are depicted in Figure 11.3. More than half of the studies (19/30) show a bias for mutation detection in the non-duplicated 3' region of the gene. Thirty-six percent (67/185) of all reported mutations are clustered in the unique region, which accounts for only ~2.5kb of the gene product (International PKD Consortium, 1995). Only 6-7% of individuals with ADPKD linked to the *PKDI* gene have detectable mutations in the unique region of the gene (see Table 11.3), this region represents approximately 18% of polycystin-1.

Mutations have been detected in each exon of the 3' region of *PKDI*. A non-disease-causing VNTR is located in IVS 42 (Peral *et al.*, 1996b). The seven alleles detected at this site range from 410 bp to 215 bp, representing an intron size variation of 325-130 bp. The intron is a complex structure and includes a 34 bp repeat that was present five times in the analysed clone (Peral *et al.*, 1996b). The variation may interfere with amplification by changing the primer binding sites, or result in an amplicon too large for the chosen mutation screening technique to be effective. Physical difficulties associated with amplification could perhaps explain why so few variants have been detected in exon 42 and IVS 42 of the gene.

Mutation screening results infer that the majority of germline mutations are located in the duplicated region (82%) of polycystin-1. As of February 2002, 118 mutations have been reported in the duplicated region from approximately 600 individuals analysed

(see Table II.3). A true mutation detection rate cannot be estimated, as most studies have failed to screen the entire gene, however a relative rate of detection can be observed by comparing the number of patients screened with the percentage of the gene covered (Table VII.1).





**Figure II.3.** Mutations of *PKD1*

### II.3.v. Recurrent Mutations in *PKD1*

There have been no reports on any common recurring pathogenic mutations in *PKD1* and no mutational hot-spots are apparent in the gene. As indicated in Table II.5 the nonsense mutation, Q4041X, in exon 44 has been reported in four families (Turco *et al.*, 1995; Daniells *et al.*, 1998b; Torra *et al.*, 1998). Comparisons of polymorphic haplotypes of the Q4041X-bearing chromosomes in families with this mutation revealed no identity by descent (Torra *et al.*, 1998). A slipped-mispairing model has been proposed to explain the recurrence of the mutation (Daniells *et al.*, 1998b). Whatever the mechanism of origin, the most common *PKD1* mutation known to date, Q4041X, accounts for only 4 of approximately 950 disease alleles.

Other possible recurring mutations that have been found include: a small deletion in exon 15 (5224del2) in three families (Watnick *et al.*, 1999); a nonsense mutation in exon 46, R4227X in two families (Peral *et al.*, 1996b & 1997); and a missense mutation, R4275W in exon 46 (Badenas *et al.*, 1999) in a further two families. Most mutation detection studies so far conducted for ADPKD have involved multi-generational families. Spontaneous (*de novo*) mutations are estimated to account for 10% or less of all reported ADPKD cases (Gabow, 1993 & 1996; Koptides *et al.*, 2000).

### II.3.vi. Mutational Effects on Protein

The majority of *PKD1* mutations detected result in a truncated protein that lacks a part of or possibly the entire C-terminus of polycystin-1, suggesting that this region has an important function (Li *et al.*, 1999). A nonsense mutation (R4227X) towards the 3' end of the gene results in the loss of a phosphorylation site in the polycystin-1 C-terminal

domain at serine residue 4251. It is proposed that this serine is the site of phosphorylation by a cAMP-dependent protein kinase in normal human renal collecting tubule cells and may be necessary for normal kidney function (Li *et al.*, 1999).

Until more ADPKD genotypes (including both truncating and non-truncating mutations) can be related to the function of *PKD1* and *PKD2* gene products, and their interactions or pathways, it will not be possible to predict the severity of particular mutations. Hence the importance of gathering phenotypic information, and using it in the correlation of mutations resulting from truncated proteins and missense changes in particular protein domains. No significant clustering of mutations, either truncating or in-frame, in any particular region of the gene have been reported in *PKD1*, however, trends such as the number of missense mutations in exons 23 and 25 (Watnick *et al.*, 1997) are emerging. The occurrence of missense mutations in these exons, and other patterns, will be discussed in Chapter VII together with the results of mutations detected in this study.

**Table II.3. Mutations Detected in *PKD1***

<b>Mutation</b>	<b>Type</b>	<b>Reference</b>
Total deletion of the <i>PKD1</i> gene	Large deletion (entire gene)	European PKD Consortium, (1994).
c.224del13	Small deletion Frameshift exon 1	Rossetti <i>et al.</i> , (2001)
L13Q	Missense exon 1	Rossetti <i>et al.</i> , (2001)
g.18177del 3 kb IVS 1-exon 5	Large deletion Frameshift	Thomas <i>et al.</i> , (1999)
S75F	Missense exon 2	Rossetti <i>et al.</i> , (2001)
W139C	Missense exon 4	Rossetti <i>et al.</i> , (2001)
S225X	Nonsense exon 5	Rossetti <i>et al.</i> , (2001)
Q227X	Nonsense exon 5	Rossetti <i>et al.</i> , (2001)
R324L	Missense exon 5	Thomas <i>et al.</i> , (1999)
IVS7+1G>A	Splicing mutation Frameshift	Rossetti <i>et al.</i> , (2001)
IVS9+1G>T	Splicing mutation Frameshift	Rossetti <i>et al.</i> , (2001)
c.2296delC	Small deletion Frameshift exon 10	Rossetti <i>et al.</i> , (2001)
c.2296insC	Small insertion Frameshift exon 10	Rossetti <i>et al.</i> , (2001)
Q705X	Nonsense exon 11	Rossetti <i>et al.</i> , (2001)
S74S	Nonsense exon 11	Rossetti <i>et al.</i> , (2001)
L845S	Missense exon 11	Thomas <i>et al.</i> , (1999)
W967R	Missense exon 12	Phakdeekitcharoen <i>et al.</i> , (2001)
c.3220del4	Small deletion Frameshift exon 13	Watnick <i>et al.</i> , (1999)
c.3336delG	Small deletion Frameshift exon 13	Phakdeekitcharoen <i>et al.</i> , (2000)
IVS13-2A>T	Splicing mutation	Thongnoppakhun <i>et al.</i> , (2000)
IVS 14 -1 G>A	Splicing mutation	Thomas <i>et al.</i> , (1999)

G1166S	Missense exon 15	Phakdeekitcharoen <i>et al.</i> , (2000)
c.3804del7	Small deletion Frameshift exon 15	Watnick <i>et al.</i> , (1999)
c.3866delT	Small deletion exon 15	Roelfsema <i>et al.</i> , (1997)
c.4036/42del2	Small deletion exon 15	Roelfsema <i>et al.</i> , (1997)
c.4137delCT	Small deletion Frameshift exon 15	Rossetti <i>et al.</i> , (2001)
c.4291delG	Small deletion Frameshift exon 15	Rossetti <i>et al.</i> , (2001)
c.4787delG	Small deletion Frameshift exon 15	Rossetti <i>et al.</i> , (2001)
E1537X	Nonsense exon 15	Rossetti <i>et al.</i> , (2001)
c.4898insT	Small insertion Frameshift exon 15	Thomas <i>et al.</i> , (1999)
c.5014delA	Small deletion exon 15	Roelfsema <i>et al.</i> , (1997)
Q1653X	Nonsense exon 15	Phakdeekitcharoen <i>et al.</i> , (2000)
c.5224del2	Small deletion Frameshift exon 15	Watnick <i>et al.</i> , (1999)
W1826X	Nonsense exon 15	Watnick <i>et al.</i> , (1999) (F1826X in paper !)
c.5782del1	Small deletion Frameshift exon 15	Watnick <i>et al.</i> , (1999b)
W1874X	Nonsense exon 15	Roelfsema <i>et al.</i> , (1997)
c.5870del14	Small deletion Frameshift exon 15	Rossetti <i>et al.</i> , (2001)
Q1922X	Nonsense exon 15	Thomas <i>et al.</i> , (1999)
c.6187delC	Small deletion In-frame exon 15	Rossetti <i>et al.</i> , (2001)
V1956E	Missense exon 15	Phakdeekitcharoen <i>et al.</i> , (2000)

Q1960X	Nonsense exon 15	Phakdeekitcharoen <i>et al.</i> , (2000)
Q2039X	Nonsense exon 15	Phakdeekitcharoen <i>et al.</i> , (2000)
c.6434del28	Large deletion exon 15	Thomas <i>et al.</i> , (1999)
c.6574del7	Small deletion exon 15	Roelfsema <i>et al.</i> , (1997)
R2163X	Nonsense exon 15	Rossetti <i>et al.</i> , (2001)
c.6868del15	Small deletion In-frame exon 15	Rossetti <i>et al.</i> , (2001)
c.6937delAC	Small deletion Frameshift exon 15	Rossetti <i>et al.</i> , (2001)
Q2243X	Nonsense exon 15	Rossetti <i>et al.</i> , (2001)
C2229X	Nonsense exon 15	Rossetti <i>et al.</i> , (2001)
T2250M	Missense exon 15	Perrichot <i>et al.</i> , (2000b)
IVS15 +2 T>G c.7174/7179del14	Splicing mutation Small deletion Frameshift exon 16	Perrichot <i>et al.</i> , (2000b) Bouba <i>et al.</i> , (2001)
R2329W	Missense exon 16	Perrichot <i>et al.</i> , (2000b)
c.7205del7	Small deletion Frameshift exon 16	Phakdeekitcharoen <i>et al.</i> , (2000)
c.7211ins7	Small insertion Frameshift exon 16	Rossetti <i>et al.</i> , (2001)
Y2336D	Missense exon 16	Rossetti <i>et al.</i> , (2001)
Y2336X	Nonsense exon 16	Perrichot <i>et al.</i> , (2000b)
7277 -3 C>G IVS 16/exon 17 C2370X	Splicing mutation Frameshift Nonsense exon 17	Watnick <i>et al.</i> , (1999) Rossetti <i>et al.</i> , (2001)
c.7324delGT	Small deletion Frameshift exon 17	Rossetti <i>et al.</i> , (2001)
c.7397ins11	Small insertion Frameshift exon 17	Watnick <i>et al.</i> , (1999)

R2402X	Nonsense exon 17	Phakdeekitcharoen <i>et al.</i> , (2000)
R2408C	Missense exon 18	Phakdeekitcharoen <i>et al.</i> , (2000)
R2430X	Nonsense exon 18	Phakdeekitcharoen <i>et al.</i> , (2000)
c.7535ins3	Small insertion In-frame exon 18	Phakdeekitcharoen <i>et al.</i> , (2000)
P2471L	Missense exon 18	Bouba <i>et al.</i> , (2001)
Q2519L	Missense exon 19	Bouba <i>et al.</i> , (2001)
Q2558X	Nonsense exon 19	Phakdeekitcharoen <i>et al.</i> , (2000)
c.7946del3	Small deletion In-frame exon 20	Bouba <i>et al.</i> , (2001)
c.8030del5	Small deletion Frameshift exon 20	Perrichot <i>et al.</i> , (2000b)
c.8047del3	Small deletion In-frame exon 20	Bouba <i>et al.</i> , (2001)
R2639X	Nonsense exon 21	Perrichot <i>et al.</i> , (2000b)
c.8126ins20	Small insertion Frameshift exon 21	Rossetti <i>et al.</i> , (2001)
T2649I	Missense exon 21	Bouba <i>et al.</i> , (2001)
c.8159delCT	Small deletion Frameshift exon 21	Phakdeekitcharoen <i>et al.</i> , (2000)
c.8183del8	Small deletion Frameshift exon 21	Bouba <i>et al.</i> , (2001)
IVS21-2delAG L2696R	Splicing mutation Missense exon 22	Rossetti <i>et al.</i> , (2001) Phakdeekitcharoen <i>et al.</i> , (2001)
A2752D	Missense exon 23	Rossetti <i>et al.</i> , (2001)
M2760T	Missense exon 23	Watnick <i>et al.</i> , (1997)
R2761P	Missense exon 23	Watnick <i>et al.</i> , (1997)
L2763V	Missense exon 23	Watnick <i>et al.</i> , (1997)

M2764T	Missense exon 23	Watnick <i>et al.</i> , (1997)
c.8507ins12	Small insertion In-frame exon 23	Rossetti <i>et al.</i> , (2001)
V2768M	Missense exon 23	Rossetti <i>et al.</i> , (2001)
E2771K	Missense exon 23	Rossetti <i>et al.</i> , (2001)
R2791Q	Missense exon 23	Watnick <i>et al.</i> , (1997)
L2816P	Missense exon 23	Rossetti <i>et al.</i> , (2001)
I2826T	Missense exon 23	Watnick <i>et al.</i> , (1997)
c.8657delC	Small deletion Frameshift exon 23	Peral <i>et al.</i> , (1997)
G2858S	Missense exon 23	Rossetti <i>et al.</i> , (2001)
Q2900X	Nonsense exon 23	Roelfsema <i>et al.</i> , (1997)
H2921P	Missense exon 23	Koptides <i>et al.</i> , (2000)
c.9142/9143del3	Small deletion In-frame exon 24	Bouba <i>et al.</i> , (2001)
E2966D	Missense exon 25	Afzal <i>et al.</i> , (2000)
R2985G	Missense exon 25	Phakdeekitcharoen <i>et al.</i> , (2001)
L2993P	Missense exon 25	Peral <i>et al.</i> , (1997)
W3001X	Nonsense exon 25	Phakdeekitcharoen <i>et al.</i> , (2001)
V3008L	Missense exon 25	Watnick <i>et al.</i> , (1997)
c.9245del18	Small deletion In-frame exon 25	Rossetti <i>et al.</i> , (2001)
Q3016R	Missense exon 25	Peral <i>et al.</i> , (1997)
E3020X	Nonsense exon 25	Peral <i>et al.</i> , (1997)
R3039C	Missense exon 25	Phakdeekitcharoen <i>et al.</i> , (2001)
F3064L	Missense exon 25	Watnick <i>et al.</i> , (1997)



c.9299delC	Small deletion Frameshift exon 25	Peral <i>et al.</i> , (1997)
IVS25-16G>A	Splicing mutation Frameshift	Rossetti <i>et al.</i> , (2001)
W3180X	Nonsense exon 27	Rossetti <i>et al.</i> , (2001)
c.9867delT	Small deletion Frameshift exon 28	Perrichot <i>et al.</i> , (2000b)
Q3206X	Nonsense exon 28	Perrichot <i>et al.</i> , (2000b)
R3247H	Missense exon 29	Afzal <i>et al.</i> , (2000)
V3285I	Missense exon 29	Phakdeekitcharoen <i>et al.</i> , (2001)
c.10135-2A>T	Splicing mutation IVS29	Afzal <i>et al.</i> , (2000)
V3375M	Missense exon 31	Koptides <i>et al.</i> , (1998)
T3382M	Missense exon 31	Afzal <i>et al.</i> , (2000)
Deletion of 2 kb exon 31-exon 34 c.10378 + 25del19 IVS31 Q3394X	Large deletion Frameshift Small deletion Frameshift Nonsense exon 32	European PKD Consortium, (1994) Peral <i>et al.</i> , (1997) Rossetti <i>et al.</i> , (2001)
Deletion of 5.5 kb of genomic DNA Q3474X	Large deletion exon 34- 3'UTR Nonsense exon 34	European PKD Consortium, (1994) Perrichot <i>et al.</i> , (1999)
T3509M*	Missense exon 35	Mizoguchi <i>et al.</i> , (2001)
L3510V	Missense exon 35	Peral <i>et al.</i> , (1997)
Q3513X	Nonsense exon 35	Peral <i>et al.</i> , (1997)
G3559R	Missense exon 36	Tsuchiya <i>et al.</i> , (2001)
c.10946insT	Small insertion exon 36	Peral <i>et al.</i> , (1996b)
E3631D	Missense exon 37	Peral <i>et al.</i> , (1996b)
P3648L	Missense exon 37	Ding <i>et al.</i> , (2002)
M3677T	Missense exon 38	Turco <i>et al.</i> , (1997)

C3693X	Nonsense exon 38	Mizoguchi <i>et al.</i> , (2001)
R3718Q	Missense exon 38	Aguiari <i>et al.</i> , (2000)
c.11284insT	Small insertion Frameshift exon 38	Perrichot <i>et al.</i> , (1999)
c.11285insC	Small insertion Frameshift exon 38	Perrichot <i>et al.</i> , (1999)
c.11307del61	Large deletion exon 38	Tsuchiya <i>et al.</i> , (2001)
11449del15 (cDNA level)	Small deletion exon 39	Afzal <i>et al.</i> , (1999)
c.11453del15	Small deletion in-frame exon 39	Peral <i>et al.</i> , (1996b)
R3752W	Missense exon 39	Kim <i>et al.</i> , (2000b)
I39E40-25 to I39E40+47 72bp deletion	Large deletion	Peral <i>et al.</i> , (1996b)
IVS39+1 G>C	Splicing mutation	Peral <i>et al.</i> , (1997)
c.11549ins10 10 bp duplication	Small insertion Frameshift exon 40	Turco <i>et al.</i> , (1997)
c.11548del8	Small deletion Frameshift exon 40	Kim <i>et al.</i> , (2000b)
C.11674insG	Small insertion Frameshift exon 41	Kim <i>et al.</i> , (2000b)
Y3818X	Nonsense exon 41	Peral <i>et al.</i> , (1996a)
Q3820X	Nonsense exon 41	Perrichot <i>et al.</i> , (2000a)
Q3837X	Nonsense exon 41	Peral <i>et al.</i> , (1996b)
D3814N	Missense exon 41	Kim <i>et al.</i> , (2000b)
IVS41 +2ins3 **	Splicing mutation	Perrichot <i>et al.</i> , (2000a)
L3851P	Missense exon 42	Aguiari <i>et al.</i> , (2000)
c.12187ins9	Small insertion exon 43	Perrichot <i>et al.</i> , (1999)
c.12714ins23	Large insertion duplication exon 43	Perrichot <i>et al.</i> , (1999)

18 bp deletion E43IVS43 +15/16	Small Deletion Splicing	Peral <i>et al.</i> , (1995)
20 bp deletion E43IVS43 + 5-13	Small Deletion Splicing	Peral <i>et al.</i> , (1995)
15 bp del in 75bp intron Paper states IVS 47!	Splicing mutation Probably IVS 43?	Tighe <i>et al.</i> , (1998)
Q4010X	Nonsense exon 44	Daniells <i>et al.</i> , (1998b)
W4011X	Nonsense exon 44	Roelfsema <i>et al.</i> , (1997)
R4020X	Nonsense exon 44	Rossetti <i>et al.</i> , (1996)
E4024X	Nonsense exon 44	Daniells <i>et al.</i> , (1998b)
G4031D	Missense exon 44	Daniells <i>et al.</i> , (1998b)
Y4039X	Nonsense exon 44	Rossetti <i>et al.</i> , (2001)
Q4041X	Nonsense exon 44	Turco <i>et al.</i> , (1995)
c.12252del2	Small deletion Frameshift exon 44	Afzal <i>et al.</i> , (1999)
c.12262delAT	Small deletion Frameshift exon 44	Daniells <i>et al.</i> , (1998b)
c.12290insG	Small insertion Frameshift exon 44	Afzal <i>et al.</i> , (1999)
IVS 44 + 1 G>C	Splicing mutation In-frame	European PKD Consortium, (1994)
IVS44 -1 G>C	Splicing mutation IVS 44	Badenas <i>et al.</i> , (1999)
Q4059X	Nonsense exon 45	Daniells <i>et al.</i> , (1998a)
C4086X	Nonsense exon 45	Neophytou <i>et al.</i> , (1996)
Q4124X	Nonsense exon 45	Daniells <i>et al.</i> , (1998a)
Y4126X	Nonsense exon 45	Turco <i>et al.</i> , (1997)
R4135G	Missense exon 45	Perrichot <i>et al.</i> , (1999)
W4139X	Nonsense exon 45	Perrichot <i>et al.</i> , (1999)
c.12412ins20	Small insertion Frameshift exon 45	Daniells <i>et al.</i> , (1998a)

c.12470insA	Small insertion Frameshift exon 45	Ding <i>et al.</i> , (2002)
c.12506insG	Small insertion Frameshift exon 45	Daniells <i>et al.</i> , (1998a)
c.12593delA	Small deletion Frameshift exon 45	Ding <i>et al.</i> , (2002)
c.12601del3	Small deletion In frame exon 45	Afzal <i>et al.</i> , (1999)
c.12617delC	Small deletion Frameshift exon 45	Rossetti <i>et al.</i> , (2001)
IVS45+55del25	Large deletion intron 45	Aguiari <i>et al.</i> , (2000)
IVS 45 -1 G>A R4153C	Splicing mutation Missense exon 46	Badenas <i>et al.</i> , (1999) Perrichot <i>et al.</i> , (1999)
Q4224P	Missense exon 46	Badenas <i>et al.</i> , (1999)
R4227X	Nonsense exon 46	Peral <i>et al.</i> , (1996b)
Y4236X	Nonsense exon 46	Perrichot <i>et al.</i> , (2000a)
12722delT (cDNA level)	Small deletion Frameshift exon 46	Kim <i>et al.</i> , (2000b)
c.12739delA	Small deletion Frameshift exon 46	Peral <i>et al.</i> , (1997)
c.12801del28	Large deletion exon 46	Torra <i>et al.</i> , (1997)
R4275W	Missense exon 46	Badenas <i>et al.</i> , (1999)

\*T3509M Mizoguchi *et al.*, (2001) list this change as disease-causing, Peral *et al.*, (1997) and the present study propose that the same variation is neutral.

\*\* IVS41 +2ins3 Perrichot *et al.*, (2000a) list this change as disease-causing, Koptides *et al.*, (2000) cite the same variation as neutral although no RNA work was performed.

### 11.3.vii. Neutral Polymorphisms in the Human Genome

Single nucleotide polymorphisms (SNPs) represent the most frequent type of variation in the human genome. It was originally estimated that they occur at about 1 per 1000 bp in a population (Wang *et al.*, 1998), but more recent estimates from the International SNP Map Working Group are 1 SNP per 1.9kb, with a higher SNP occurrence within genes (Sachidanandam *et al.*, 2001). The genetic diversity created by SNPs could contribute to the basis of heritable variation in disease susceptibility and supply information on human migrations (Wang *et al.*, 1998). There has been increased appreciation that large collections of SNPs could provide a useful tool for studies into the genetic basis of complex human diseases (Risch & Merikangas, 1996). For whole-genome association studies, it has been calculated that useful levels of linkage disequilibrium are unlikely to extend beyond a distance of 3 kb in the general population, and so for screening to be informative approximately 500,000 appropriately spaced SNPs are required (Kruglyak, 1999).

In practice, one SNP for a particular gene may be informative and associated with a disease gene, whilst another in the same gene may well be informative as a genetic marker, however, not be associated with disease susceptibility. For example, there are tens of polymorphisms found in the beta-globin gene but only one single nucleotide polymorphism in this gene is associated with sickle-cell disease (Roses, 1997). Two particular SNPs within the sequence of apolipoprotein E are phenotype-associated; one is responsible for the transition from APOE4 to APOE3, and the other from APOE3 to APOE2. APOE4 is associated with an increased risk and lower age at onset of Alzheimer's disease, whereas APOE2 is associated with decreased risk and older age at

onset of Alzheimer's disease (Roses, 1997). When choosing SNPs for genome screening, it will be necessary not only to monitor how they are spaced as informative genetic markers, but also to consider their biological relevance to the disease (Roses, 1997).

Obviously it is not only work specific to SNP discovery which will record polymorphisms in the genome. All mutation detection screening studies, including the present investigation, have the potential to uncover disease-causing and neutral polymorphisms. The HGBASE database, Human Genic Bi-Allelic Sequences database (Uppsala University & Interactiva Biotechnologie GmbH. <http://hgbase.interactiva.de/>) has the capacity to list known polymorphisms within particular human genes (Brookes *et al.*, 2000). As of February 2002 only 11 SNPs have been listed in HGBASE for the *PKD1* gene, although many more are reported in the literature (Tables II.4(a) & II.4(b)).

#### 11.3.viii. Neutral Polymorphisms in *PKD1*

Of the 132 neutral polymorphisms detected from exons 2 to the 3' UTR in the *PKD1* gene, 126 are SNPs, the remaining six consist of three small deletions and three small insertions (Table II.4.b). A number of studies have focused on exons 5 to 15; exons 17 to 21; and exon 23 to the 3' UTR, equivalent to an overall genomic area of 27,073 bp. This indicates a frequency in that portion of *PKD1* of 1 SNP per 237 bp (114 in all), a figure that is in accordance with currently reported frequencies for the rest of the genome (<http://hgbase.cgr.ki.se/>). If the proposed pathogenic SNPs of *PKD1* are added (92 in all, not normally defined as SNPs because they are disease-causing), this figure is 1 nucleotide variant per 131 bp in *PKD1*, confirming the high mutability of the gene.

The non-disease-causing polymorphisms so far recorded in the *PKD1* gene are listed in Tables II.4.a & II.4.b. Mapping and understanding this high level of polymorphism in the *PKD1* gene (and in the *PKD2* gene), could lead to a better comprehension of the mutation rates and mutational mechanisms associated with ADPKD. The presence or absence of particular non-disease-causing polymorphisms may also correlate with clinical heterogeneity in ADPKD (Watnick *et al.*, 1999). Some neutral polymorphisms may associate with particular disease-causing mutations displaying a similar phenotype, or they may influence phenotypic heterogeneity between family members with identical germline mutations. Although for diagnostic purposes it is the detection of disease-causing mutations that is of primary concern, all neutral polymorphisms should be noted for future consideration with regard to clinical heterogeneity and association with particular disease-causing mutations.

**Table II.4(a). Non-disease Causing Missense Changes Detected in *PKD1***

<b>Neutral Missense Change</b>	<b>Location</b>	<b>Reference</b>
4406T/A - W/R1399	Exon 15	Thomas <i>et al.</i> , (1999)
4406T/C - W/R1399	Exon 15	Watnick <i>et al.</i> , (1999)
5568C/T - P/L1786	Exon 15	Thomas <i>et al.</i> , (1999)
6195G/A - R/H1995	Exon 15	Phakdeekitcharoen <i>et al.</i> , (2000)
7347A/G - Y/C2379	Exon 17	Perrichot <i>et al.</i> , (2000b)
7386G/C - R/P2392	Exon 17	Watnick <i>et al.</i> , (1999) ##
7479C/T - S/F2423	Exon 18	Watnick <i>et al.</i> , (1999) ##
7853G/C - E/Q2548	Exon 19	Watnick <i>et al.</i> , (1999)
7956C/T - T/M2582	Exon 20	Bouba <i>et al.</i> , (2001)
8021G/A - D/N2604	Exon 20	Phakdeekitcharoen <i>et al.</i> , (2000)
8124A/G - H/R2638	Exon 21	Watnick <i>et al.</i> , (1999)
8231C/T - P/S2674	Exon 22	Rossetti <i>et al.</i> , (2001)
8334C/T - T/M2708	Exon 22	Rossetti <i>et al.</i> , (2001)
8411C/A - P/T2734	Exon 23	Rossetti <i>et al.</i> , (2001)
8415A/T - Q/L2735	Exon 23	Rossetti <i>et al.</i> , (2001)
8504C/T - R/C2765	Exon 23	Rossetti <i>et al.</i> , (2001)
8556G/A - V/M2782	Exon 23	Rossetti <i>et al.</i> , (2001)
8651G/A - G/R2814	Exon 23	Rossetti <i>et al.</i> , (2001)
8873C/G - R/G2888	Exon 23	Rossetti <i>et al.</i> , (2001)
8924G/A - V/I2905	Exon 23	Rossetti <i>et al.</i> , (2001)
9019G/C - E/D2966	Exon 24	Rossetti <i>et al.</i> , (2001)
9125G/A - D/N2972	Exon 24	Bouba <i>et al.</i> , (2001)
9235G/A - V/M3008	Exon 25	Mizoguchi <i>et al.</i> , (2001)
9406GT/CC - F/L3066	Exon 25	Peral <i>et al.</i> , (1997)
9407T/C - F/L3066	Exon 25	Perrichot <i>et al.</i> , (2000b)
9627G/T - G/V3139	Exon 27	Perrichot <i>et al.</i> , (2000b)
9789C/T - P/L3193	Exon 28	Perrichot <i>et al.</i> , (2000b)
10737C/T - T/M3509	Exon 35	Peral <i>et al.</i> , (1997).
10743C/T - A/V3511	Exon 35	Peral <i>et al.</i> , (1997).
10893G/A - S/N3561	Exon 36	Perrichot <i>et al.</i> , (1999)
12168C/G - A/G3985	Exon 43	Badenas <i>et al.</i> , (1999)
12341A/G - I/V4044	Exon 44	Rossetti <i>et al.</i> , (1996)
12384C/T - A/V4058	Exon 45	Constantinides <i>et al.</i> , (1997)
12777C/T - S/F4189	Exon 46	Peral <i>et al.</i> , (1997)

## Watnick *et al.*, (1999). Only found in affected members with aneurysm (may be a polymorphism associated with disease phenotype).



**Table II.4(b). Neutral Polymorphisms Detected in *PKD1***

<b>Polymorphism</b>	<b>Location</b>	<b>Reference</b>
487G/A – A92	Exon 2	Rossetti <i>et al.</i> , (2001)
1023C/T (1234C/T) - A341*	Exon 5	Thomas <i>et al.</i> , (1999)
1330T/C – L373	Exon 5	Rossetti <i>et al.</i> , (2001)
1420C/T – H403	Exon 6	Rossetti <i>et al.</i> , (2001)
1921C/T – H570	Exon 7	Rossetti <i>et al.</i> , (2001)
2694A/C (2905A/C) - A898*	Exon11	Thomas <i>et al.</i> , (1999)
2700G/A (2911G/A) - P900*	Exon 11	Thomas <i>et al.</i> , (1999)
2730C/T (2941C/T) - D910*	Exon 11	Thomas <i>et al.</i> , (1999)
2911G/A - P900	Exon 11	Watnick <i>et al.</i> , (1999)
2941C/T - D910	Exon 11	Watnick <i>et al.</i> , (1999)
g.25134C/T	IVS 11	Thomas <i>et al.</i> , (1999)
2972T/G (3183T/G) - V991*	Exon 12	Thomas <i>et al.</i> , (1999)
3274T/C - G1021	Exon13	Watnick <i>et al.</i> , (1999)
3111A/G (3322A/G) - L1037*	Exon13	Thomas <i>et al.</i> , (1999).
3372C/T (3583C/T) - A1124*	Exon 15	Thomas <i>et al.</i> , (1999)
3375C/T (3586C/T) - S1125*	Exon 15	Thomas <i>et al.</i> , (1999)
3864C/T (4075C/T) - H1288*	Exon 15	Thomas <i>et al.</i> , (1999)
4635G/A (4846G/A) - K1545*	Exon 15	Thomas <i>et al.</i> , (1999)
4876A/C - A1555	Exon 15	Watnick <i>et al.</i> , (1999)
4885G/A - T1558	Exon 15	Watnick <i>et al.</i> , (1999)
5383C/T - T1724	Exon 15	Watnick <i>et al.</i> , (1999)
5359C/T (5570C/T) - L1787*	Exon 15	Thomas <i>et al.</i> , (1999)
5485C/T (5696C/T) - L1829*	Exon 15	Thomas <i>et al.</i> , (1999)
5763G/A (5974G/A) - L1921*	Exon 15	Thomas <i>et al.</i> , (1999)
6058C/T – S1949	Exon 15	Phakdeekitcharoen <i>et al.</i> , (2000)
7138C/T - G2309	Exon 16	Perrichot <i>et al.</i> , (2000b)
7147G/A – A2312	Exon 16	Rossetti <i>et al.</i> , (2001)
IVS 16 –9 G/A	IVS 16	Perrichot <i>et al.</i> , (2000b)
7345G/T - V2378	Exon 17	Watnick <i>et al.</i> , (1999) ##
7376T/C - L2389	Exon17	Watnick <i>et al.</i> , (1999)
IVS 17 – 27 C/A	IVS 17	Perrichot <i>et al.</i> , (2000b)
IVS 17 – 10 C/A	IVS 17	Perrichot <i>et al.</i> , (2000b)
7652C/T - L2472 **	Exon 18	Watnick <i>et al.</i> , (1999)
7652C/T - L2481	Exon 18	Perrichot <i>et al.</i> , (2000b)
7696C/T – C2495	Exon 18	Phakdeekitcharoen <i>et al.</i> , (2000)
7919T/C - L2570	Exon 20	Watnick <i>et al.</i> , (1999)
8002A/G – P2597	Exon 20	Bouba <i>et al.</i> , (2001)
8446T/G - S2745	Exon 23	Watnick <i>et al.</i> , (1997)
8650C/T – S2813	Exon 23	Rossetti <i>et al.</i> , (2001)
8890C/G – S2893	Exon 23	Rossetti <i>et al.</i> , (2001)
9124T/C - A2971	Exon 24	Watnick <i>et al.</i> , (1997)
9159 + 14A/G	IVS 24	Watnick <i>et al.</i> , (1997)
IVS 24 – 17 A/G	IVS 24	Perrichot <i>et al.</i> , (2000b)

9175G/A - A2988	Exon 25	Perrichot <i>et al.</i> , (2000b)
g.38997C/A - L2995	Exon 25	Mizoguchi <i>et al.</i> , (2001)
9406G/C - V3065	Exon 25	Watnick <i>et al.</i> , (1997)
9481C/T - V3090	Exon 26	Perrichot <i>et al.</i> , (2000b)
9541T/C - P3110	Exon 26	Peral <i>et al.</i> , (1997)
9880G/A - T3223	Exon 28	Peral <i>et al.</i> , (1997)
IVS 31 +7 A/G	IVS 31	Perrichot <i>et al.</i> , (2000b)
IVS 34 +20 G/A	IVS 34	Perrichot <i>et al.</i> , (2000b)
9481C/T - V3090	Exon 36	Koptides <i>et al.</i> , (2000)
10765C/T - L3589* (10976C/T)	Exon 36	Aguiari <i>et al.</i> , (2000)
10976C/A - L3589 *** is this L3589M or 10976C/T L3589	Exon 36	Perrichot <i>et al.</i> , (1999)
g.47848C/T	IVS 36	Mizoguchi <i>et al.</i> , (2001)
g.47853delC	IVS 36	Mizoguchi <i>et al.</i> , (2001)
10869G/A P3623* (11080G/A)	Exon 37	Aguiari <i>et al.</i> , (2000)
g.48085C/T	IVS 37	Mizoguchi <i>et al.</i> , (2001)
g.48205C/G	IVS 37	Mizoguchi <i>et al.</i> , (2001)
11364+ 13 G>A	IVS 38	Perrichot <i>et al.</i> , (1999)
11477 +17del14	IVS 39	Badenas <i>et al.</i> , (1999)
11521G/A - A3371	Exon 40	Rossetti <i>et al.</i> , (2001)
g.48509G/A - L3674	Exon 38	Mizoguchi <i>et al.</i> , (2001)
L3753	Exon 39	Tsuchiya <i>et al.</i> , (2001)
11584G/C - S3791	Exon 40	Peral <i>et al.</i> , (1996b)
g.49637G/A	IVS 40	Mizoguchi <i>et al.</i> , (2001)
IVS41 +2ins3 ***	IVS 41	Koptides <i>et al.</i> , (2000)
IVS41 +8ins3	IVS 41	Aguiari <i>et al.</i> , (2000)
g.49833C/T	IVS 41	Mizoguchi <i>et al.</i> , (2001)
g.49942T/G	IVS 41	Mizoguchi <i>et al.</i> , (2001)
g.49942T/G	IVS 41	Mizoguchi <i>et al.</i> , (2001)
g.50075G/A - E3871	Exon 42	Mizoguchi <i>et al.</i> , (2001)
g.50493C/A - A3910	Exon 43	Mizoguchi <i>et al.</i> , (2001)
12124C/T - R3971	Exon 43	Perrichot <i>et al.</i> , (1999)
12184C/G - A3991	Exon 43	Badenas <i>et al.</i> , (1999)
12307G/C - L4032	Exon 43	Badenas <i>et al.</i> , (1999)
12212 - 34C/A	IVS 43	Daniells <i>et al.</i> , (1998b)
12211 + 37ins5	IVS 43	Badenas <i>et al.</i> , (1999)
12346 +19delG	IVS 44	Daniells <i>et al.</i> , (1998b)
12485C/G - A4091*** (L4092V)	Exon 45	Perrichot <i>et al.</i> , (1999)
or 12483C/G - A4091G		
12271A>T/G/C - A4091* (12484A>T/G/C)	Exon 45	Aguiari <i>et al.</i> , (2000)
12617C/T - L4136	Exon 45	Badenas <i>et al.</i> , (1999)

12652+ 34 G>T	IVS 45	Perrichot <i>et al.</i> , (1999)
12838T/C - P4209	Exon 46	Peral <i>et al.</i> , (1997)
<b>g.51659C/T</b> - D4234	Exon 46	Mizoguchi <i>et al.</i> , (2001)
<b>12444C/T</b> - P4254* (12655C/T)	Exon 46	Aguiari <i>et al.</i> , (2000)
12973C/T - P4254	Exon 46	Peral <i>et al.</i> , (1996b)
13133G>A	3' UTR	Perrichot <i>et al.</i> , (1999)
13135G/A	3' UTR	Badenas <i>et al.</i> , (1999).
13233A/G	3' UTR	Badenas <i>et al.</i> , (1999)
13357T/C	3' UTR	Badenas <i>et al.</i> , (1999)
13364C/T	3' UTR	Badenas <i>et al.</i> , (1999)
13370G/A	3' UTR	Badenas <i>et al.</i> , (1999)

\*Both Thomas *et al.*, (1999) and Aguiari *et al.*, (2000) published nucleotide positions (in bold) that do not correspond to the c.DNA sequence (Acc. L33243), the nucleotide positions that do correspond to L33243 are given in brackets.

\*\* Watnick *et al.*, (1999). If 7652C/T is correct then amino acid should read L248I. If not the c.DNA position maybe 7625C/T.

\*\*\* Perrichot *et al.*, (1999) published 2 polymorphisms 10976C/A and 12485C/G where the corresponding amino acid change is unclear.

\*\*\*\* Koptides *et al.*, (2000) list IVS41 +2ins3 as neutral, Perrichot *et al.*, (2000a), cite the same variation as disease-causing, no RNA work was performed.

## II.4. Genotype-Phenotype Correlation in Single-Gene Disorders

### II.4.i. Phenotypic Variation

A topic of particular interest is why some families, or individuals within families, are severely affected by a monogenic inherited disease whilst others are not. It is important to understand what is happening at the molecular level in order to ascertain if any prediction of disease outcome is possible and, if so, how to establish a course of treatment which may prevent, delay or minimise the onset of particular symptoms. Understanding the mode of action of specific mutations will also contribute towards a more coherent picture of gene structure and protein function (Waters *et al.*, 1998). The prevalence of ADPKD in the general Japanese population has been estimated at 1 in 4,000, approximately one quarter of that in Caucasians (Higashihara *et al.*, 1998), whereas the prevalence diagnosed at autopsy in Japan is about 1 in 400 (Higashihara *et al.*, 1998). This discrepancy suggests that a large proportion of ADPKD individuals do not progress to ESRF during their lifetimes (Higashihara *et al.*, 1998). The less severe clinical phenotype does not correspond to fewer families linked to *PKD1* (Wang *et al.*, 1995), so there must be other mechanisms, apart from locus heterogeneity, which control phenotypic variation in this population. In addition to the genotype associated with a disease gene, other modifying factors, both genetic and environmental, must be considered when comparing disease phenotypes (Dipple & McCabe, 2000).

#### II.4.ii. Genotype-Phenotype Correlation Studies

Genotype-phenotype correlations have been investigated in many Mendelian diseases. Examples include phenylketonuria (PKU) (Ledley, 1991; Scriver *et al.*, 1996; Waters *et al.*, 1998); cystic fibrosis (The Cystic Fibrosis Genotype-Phenotype Consortium, 1993; Zielenski & Tsui, 1995; Ferrari & Cremonesi, 1996); adrenoleukodystrophy (Dodd *et al.*, 1997; Smith *et al.*, 1999); spinal muscular atrophies (Brahe & Bertini, 1996); inherited peripheral neuropathies (Nelis *et al.*, 1999); breast cancers associated with mutations in the *BRCA1* and *BRCA2* genes (Weber, 1998; Presneau *et al.*, 1998), and Marfan Syndrome (Ramirez *et al.*, 1999). Genotype-phenotype correlations have not yet been reported for *PKD1*. Any significant trends in phenotype variation, associated with particular types of mutations and located in regions of the gene corresponding to various protein domains, will be discussed in Chapter VII. Possible genetic factors contributing to phenotypic expression include locus heterogeneity, mutation heterogeneity (including complexities such as compound heterozygosity, the presence of double mutant alleles and neutral variants), and the genotypes of modifying genes. Examples of how each of these factors may affect the disease phenotype are outlined below.

##### II.4.ii.(a). Locus Heterogeneity

There are numerous conditions that are classified clinically as one disorder but which are in fact caused by mutations in different genes. A few examples of such conditions and their associated genes are tuberous sclerosis (*TSC1* or *TSC2*), inherited breast cancer (*BRCA1* or *BRCA2*), and hereditary non-polyposis colorectal cancer (*hMSH6*, *PMS2*, *hMLH1*, *hMSH2* or *TGFBR2*). One condition displaying a particularly high

level of locus heterogeneity is retinitis pigmentosa, with 14 loci so far associated with the clinical phenotype (<http://www.ncbi.nlm.nih.gov/OMIM/>).

As previously stated, *PKD2* is generally associated with a milder phenotype than *PKD1*. Ravine *et al.* (1992), found that non-*PKD1* patients lived longer than *PKD1* patients, had a lower risk of hypertension and progression to ESRF, and had fewer renal cysts at the time of diagnosis. A more recent multi-centre study reviewed the differences between the two phenotypes by comparing 333 people with ADPKD linked to the *PKD1* gene (33 families) and 291 people with the *PKD2*-linked disease (31 families) with 398 matched controls (Hateboer *et al.*, 1999b). It found that the median age at death or at ESRF was 53 years in *PKD1*-affected individuals, 69 years in *PKD2*-affected individuals and 78 years in normal controls (Hateboer *et al.*, 1999b).

#### II.4.ii.(b). Mutation Heterogeneity

The position and type of mutation within a gene may be used to predict its effect on gene expression. A missense variation that is not predicted to alter protein structure would not be expected to be associated with a severe phenotype, whereas a large deletion or early truncation resulting in loss of function could have severe consequences. For example, the first mutation detected in ADPKD was a deletion of most, if not all, of the *PKD1* gene (European PKD Consortium, 1994). This type of deletion is associated with very early onset ADPKD and a particularly severe phenotype (Brook-Carter *et al.*, 1994; Longa *et al.*, 1997; Harris, 1997). The *HPA* (hyperphenylalaninaemia) gene has a large number of mutations similar to *PKD1*. More than 300 mutations have been identified in *HPA* and the associated phenotype

ranges from classical PKU, with impaired cognitive development, to non-PKU HPA, with a low risk of mental retardation (Waters *et al.*, 1998).

Complex genotype-phenotype correlations exist in cystic fibrosis, with over 900 mutations found in the *CFTR* gene (<http://www.genet.sickkids.on.ca/cftr/>). While there is no obvious correlation between the genotype and the severity of most CF symptoms (The CF Genotype-Phenotype Consortium, 1993), there is a general correlation with respect to long-term pancreatic sufficiency involving a group of mutations (The CF Genotype-Phenotype Consortium, 1993). The *CFTR* model was first proposed by Riordan *et al.*, (1989) and consists of two portions each with similar nucleotide binding folds (NBD1 & 2) that connect via a regulatory domain, plus a membrane-spanning domain which forms a channel pore. CF patients who display pancreatic sufficiency generally have mutations which are found outside the NBD1 domain, whereas the most severe phenotypes are often associated with a cluster of mutations in the NBD1 domain (Nousia-Arvanitakis, 1999).

The emerging mutational pattern in ADPKD is that of a large number of private mutations confined to individual families, thus complicating the identification of molecular defects and the study of genotype-phenotype correlations. An additional factor that will impact on the feasibility of genotype-phenotype correlations is the lack of clinical detail in many published mutation studies (Daniells *et al.*, 1998a & 1998b; Roelfsema *et al.*, 1997; Thomas *et al.*, 1999; Afzal *et al.*, 1999; Thongnoppakhun *et al.*, 1999 & 2000; Rossetti *et al.*, 2001), and in the mutation database (Krawczak & Cooper, 1997).

No specific mutation has been associated with a particular phenotype, however, large deletions of the *PKD1* gene with involvement of the *TSC2* gene have shown a pattern of severe early onset renal disease (Harris, 1997). Another example of phenotypic severity associated with a particular mutation has been reported in three unrelated families with the same deletion (c.5224del2) in exon 15 of *PKD1* (Watnick *et al.*, 1999). The three families formed part of a group of 35 families who were chosen for the study because they each had some individuals with either very early onset of disease symptoms or cerebral aneurysms. The biased selection of the study group does not allow the conclusion that c.5224del2 in *PKD1* is invariably associated with a phenotype associated with early disease onset or cerebral involvement.

The most frequently occurring mutation in the *PKD1* gene is a nonsense mutation in exon 44 (Q4041X), which has been detected in four families to date (Turco *et al.*, 1995; Torra *et al.*, 1998; Daniells *et al.*, 1998b). The number of affected people in the families are, however, too small to permit a significant genotype-phenotype comparison. The phenotype details associated with Q4041X are outlined in Table II.5.



**Table 11.5.** Phenotype associated with Q4041X

Reference	Family details	Ethnicity	Sex and age	Age at CRF/ESRF/death	Other symptoms
Turco <i>et al.</i> , (1995).	VR4001  5 affected members	Italian	#1 female deceased	ESRF 56 Death 68	Hepatic cysts
			#2 male 39 years	CRF 39	Hypertension
			#3 female 72 years	ESRF 70	Hepatic cysts
			#4 male 48 years	ESRF 42 Transplant 45	Hepatic cysts
			#5 female 34 years	Normal renal function	Hepatic cysts, hypertension
Torra <i>et al.</i> , (1998) (This includes Ox 1433 detected by Peral <i>et al.</i> , 1997).	PR43  3 affected members	Spanish	#1 male 39 years	CRF (n/a)	Hepatic cysts, hypertension
			#2 male deceased	ESRF (n/a) death age 42	(n/a)
			#3 female 42 years	CRF (n/a)	(n/a)
	Ox 1433 (1 member adopted, no children)	British	Male age 60 years	CRF (n/a)	Hypertension
Daniells <i>et al.</i> , (1998b)	Family 69 No details	Caucasian (Welsh)	(n/a)	(n/a)	(n/a)
Rossetti <i>et al.</i> , (2001)	(n/a)	(n/a)	(n/a)	(n/a)	(n/a)

(n/a – no details available, age at death given is due to ESRF).

#### 11.4.ii.(c). Compound Heterozygotes and Double Mutant Alleles

Many individuals affected by single-gene disorders carry two different mutations, one on each homologous chromosome. Such individuals are termed compound heterozygotes. It has not been possible to make genotype-phenotype correlations for PKU compound heterozygotes, as the phenotype variation is too complex (Scriver, 1995).

A further complication can occur when a double mutant allele is present, with two mutations on the same chromosome. This can lead to three (or theoretically even four) different mutations in the same gene in one individual contributing to the phenotype in conjunction with other modifying factors. A number of examples have been shown, including cystic fibrosis (Savov *et al.*, 1995; Steffann *et al.*, 1998; Fanen *et al.*, 1999), homocystinuria (de Franchis *et al.*, 1999), and Marfan syndrome (Wang *et al.*, 1996b).

As ADPKD is inherited in a dominant manner, it seems improbable that a large number of compound heterozygotes could exist. Mutation detection of cystic epithelial cells has, however, revealed somatic as well as germline mutations present in both the *PKD1* and *PKD2* genes. The mutational mechanism associated with this is the “two-hit model” (Knudson, 1971) (see section 11.3.i). The phenotypic variability in ADPKD could be partially attributable to the two-hit model of mutation, with a combination of a germline mutation and variable somatic mutations. This is more complex than the phenotypic variability associated with different compound heterozygotes for PKU mutations (Scriver, 1995) as variable somatic mutations are found in the same individual.

#### II.4.ii.(d). Neutral Variants

Neutral variants are non-disease-causing polymorphisms which occur in the disease gene of interest. Whilst they alone do not appear to cause the disease, it is not yet known how or if different combinations of variants could affect phenotypic severity. It should be remembered that many genes harbour an array of polymorphisms. The PAH (Phenylalanine hydroxylase) locus (100 kb) associated with PKU has at least 18 single nucleotide polymorphisms (SNPs), eight bi-allelic restriction fragment length polymorphisms (RFLPs), and two multi-allelic sites (Scriver & Waters, 1999). The individual polymorphisms and their corresponding haplotype structures, along with allelic variation in the 5' untranslated region of the gene, may influence gene transcription and expression (Scriver & Waters, 1999).

No conclusion has been reached as to how neutral variation may (or may not) influence the phenotype of ADPKD. However, a paper by Watnick *et al.* (1999) proposed that the combined effect of a number of apparently “neutral” missense substitutions could result in a pathogenic “hypermorphic” allele. For example, mutations in the *SOD1* gene are usually associated with the autosomal dominant condition familial amyotrophic lateral sclerosis (FALS). There is a polymorphism in this gene (D90A) which, in the heterozygous state, produces no clinical effect, but homozygotes for the change display the clinical phenotype (Andersen *et al.*, 1995). It will be important to record all polymorphisms detected in the *PKD1* gene in a designated database in order to test this hypothesis in the future. Even if the hypermorphic allele is not disease-causing, the polymorphic variation may act as a modifying factor. Thus it would be advantageous

for future studies if such a database contained detailed phenotypic information along with specific genotype, *i.e.* both the mutations and the non-disease causing polymorphisms found in the *PKD1* gene.

#### II.4.ii.(e). Modifying Genes

The specific genotype of a disease gene is not the only factor which influences gene expression and phenotype severity. Gene interactions can be expected to have a functional effect. For example, phenotypic severity in CF has been difficult to correlate with the *CFTR* genotype. In a *CFTR* murine model ( $cfr^{m1HSC}/cfr^{m1HSC}$ ), little lung involvement is displayed but an intestinal obstruction similar to the meconium ileus observed in human infants with CF is present (Rozmahel *et al.*, 1996). A possible modifying gene has been reported in these mice, localised close to the centromere on mouse chromosome 7, where several genes are possible candidates for the modulation of disease severity (Rozmahel *et al.*, 1996). A region of synteny exists between this region and human chromosome 19q13 (Rozmahel *et al.*, 1996), which was demonstrated to harbour a similar human CF modifier gene (Zielenski *et al.*, 1999).

Several possible modifying genes involved in ADPKD have been examined, including the genes found in the renin-angiotensinogen-aldosterone system (RAS) proposed as the underlying cause of hypertension in ADPKD (Chapman *et al.*, 1990). A modifying factor of particular interest is the insertion/deletion polymorphism at the locus for the angiotensin I-converting enzyme (ACE) polymorphism, associated with cardiac pathology such as myocardial infarction and angina (Bøhn *et al.*, 1993; Caulfield *et al.*, 1994; Nakai *et al.*, 1994), and left ventricular hypertrophy (Schunkert *et al.*, 1994;

Prasad *et al.*, 1994). The deletion allele of this polymorphism has been implicated in an earlier age at onset of ESRF for individuals with diabetic nephropathy (both IDDM and NIDDM), IgA nephropathy (Doria *et al.*, 1994; Yoshida *et al.*, 1995), and nephroangiosclerosis (Mallamaci *et al.*, 2000). Three independent studies on the ACE polymorphism and ADPKD have suggested that individuals homozygous for the deletion allele (D/D) are at increased risk of early onset of ESRF compared to their I/D counterparts (Baboolal *et al.*, 1997; Pérez-Oller *et al.*, 1999; Van Dijk *et al.*, 1999). The poor prognosis was not invariably associated with a higher tendency towards the development of hypertension, however, an association between the D/D genotype and the occurrence of microalbuminuria has been reported (Van Dijk *et al.*, 1999). A recent study on ADPKD individuals from Australia, Bulgaria and Poland found no relationship between ADPKD severity and ACE in terms of plasma enzyme activity and I/D genotypes (Schiavello *et al.*, 2001).

#### II.4.iii. Environmental Factors

Environmental factors need to be considered when assessing the genetic contribution to the phenotype. For example, in the PiZZ form of  $\alpha$ -1 antitrypsin deficiency some of the phenotypic variation with respect to lung involvement reflects smoking behaviour (Scriver & Waters, 1999). Another major environmental modifying factor to be considered is diet. In some diseases, such as PKU or hypergalactosemia, the exclusion of certain foods (containing phenylalanine or galactose respectively) from the diet prevents the onset of clinical symptoms. As illustrated in  $\alpha$ -1 antitrypsin deficiency, life style choices and exposure to toxins and carcinogens can also contribute to phenotype. In the case of Gorlin's syndrome, a hereditary predisposition to nevoid

basal cell carcinomas (Gorlin, 1995), levels of UV radiation due to sun exposure may impact on the severity of the skin cancer phenotype.

The effect of dietary intake of protein on the rate of decline of glomerular filtration rate was investigated in a study population of 200 participants with ADPKD (Klahr *et al.*, 1995). Specifically, protein was restricted in the diet and blood pressure was controlled with the use of anti-hypertensive medication. Lower protein intake, but not prescription of a keto acid-amino acid supplement, was marginally associated with a slower progression of renal disease (Klahr *et al.*, 1995). ADPKD patients often consider a low-protein diet to delay the onset of renal failure (Maroni, 1998; Zarazaga *et al.*, 2001), however, this is not mandatory and depends on advice from clinicians. A recent study by Davis *et al.* (2001), suggests that the use of tyrosine kinase inhibitors combined with future gene therapy techniques may prove more effective than low protein diets and traditional anti-hypertensive medication.

#### II.4.vii. The importance of Genotype-Phenotype Studies for ADPKD

Genotype-phenotype correlations for ADPKD will not be possible with any degree of accuracy until many mutations have been detected. All new mutations, along with the respective phenotypes, give further information as to how truncating mutations and missense changes relate to protein function and subsequent pathogenesis. Where intra-familial variation is observed, the study of modifying factors will lead to a better understanding of how the PKD genes interact with other genes in the same pathway in the control of normal epithelial cell function, and how environmental factors can influence cystic proliferation. If genotype-phenotype correlations are established, it will

be possible to more reliably predict disease outcome and perhaps to establish if a certain course of treatment could delay the onset of particular symptoms for some patients, *i.e.* getting the right drugs to the right people.

### **III. Methodology**



### **III. Methodology (Part I)**

#### **III.1 Literature Review of Methodology**

Many screening techniques are available for mutation detection and all rely on the basic premise that they should be simple, accurate, time-efficient, economical and sensitive.

The choice of mutation detection method depends on a number of factors. The necessary equipment must be available and many techniques require a high degree of laboratory skill to yield maximal detection rates. In addition, the nature of the gene screened must also be considered. If mutations in the gene of interest are predominantly truncating, as seen in the dystrophin gene (Krawczak & Cooper, 1997), a method such as the protein truncation test (PTT) would be suitable. In contrast, if a significant proportion of the mutations are missense, as found in CFTR (Krawczak & Cooper, 1997), the conventional PTT would then not be applicable. The size of the template and its relative GC content must also be taken into account. Single-strand conformation polymorphism (SSCP) analysis as a mutation screening method has been shown to be cheap and efficient but only when PCR fragments are of optimal length (200 to 400 bp) (Orita *et al.*, 1989; Savov *et al.*, 1992).

Moreover, the quality and specificity of template influences the efficiency of all mutation detection screening techniques and sequencing methods. In the case of *PKD1* this is particularly important when trying to screen for mutations in the 5' duplicated region of the gene.

### III.1.i. Pre-Screening Methods

Prior to screening, most techniques require the amplification of gene products using the polymerase chain reaction (PCR) (Saiki *et al.*, 1986). The nature of the *PKD1* gene makes it technically challenging to amplify because of a high G-C content and the co-amplification of homologous gene transcripts. Strategies employed to produce a *PKD1*-specific template for mutation detection of the duplicated region are discussed here and novel strategies are presented in section III.5.

General techniques to improve the specificity of PCR include the addition of PCR enhancers (Kramer & Donald, 1999). In addition to the basic PCR mix, enzyme-stabilising proteins including BSA (bovine serum albumin) or gelatine, and enzyme-stabilising solvents *e.g.* glycerol, are routinely included (Kramer & Donald, 1999). Enzyme-stabilising solutes *e.g.* betaine, and solvents such as dimethyl sulphoxide (DMSO) have been shown to dramatically reduce non-specific priming events by decreasing inter- and intra-strand reannealing (Winship, 1989). Tetramethylammonium chloride (TMAC) has also been used as a PCR enhancer (Hung *et al.*, 1990). Commercially available PCR additives such as Q-solution (QIAGEN) and the Advantage-GC™ PCR Kit (Clontech) can increase reaction specificity.

### III.1.ii. DNA Screening Techniques.

#### III.1.ii.(a). Single-Strand Conformation Polymorphism Analysis

Single-strand conformation polymorphism (SSCP) analysis detects single base substitutions. It relies on the assumption that changes in nucleotide sequence affect the conformation of PCR product when it has been converted into its single-strand forms by heating and then snap-chilling the PCR products in formamide- or DMSO- based

loading buffers. The single-strand DNA is fractionated through a polyacrylamide gel under non-denaturing conditions, hence its altered electrophoretic mobility (Orita *et al.*, 1989).

This is one of the techniques most commonly used for mutation screening since it is a relatively simple procedure and gives rapid results. The length of PCR fragments that can be analysed by SSCP does, however, restrict the usefulness of the method. It was initially believed that SSCP was limited to testing PCR products of less than 200 bp (Orita *et al.*, 1989). Subsequently, it has been reported that, by increasing the percentage of the acrylamide gel, SSCP can screen PCR products greater than 400 bp in length (Savov *et al.*, 1992). Even with this increased length, the technique can be restrictive when large genes are involved, due to the numerous products that need to be analysed. The sensitivity of the method in detecting mutations must also be considered; in a blind trial on the CFTR gene it has been estimated to be as high as 84% (Jordanova *et al.*, 1997). Methods to improve the sensitivity of SSCP include using a variety of cross-linking ratios, by altering the acrylamide:bisacrylamide ratio of the gel. In addition to conventional polyacrylamide gel, a number of commercial gels (such as the MDE gel solution, BMA, Rockland, ME) are also available. The use of 5-10% glycerol in the running buffer has also helped to increase SSCP sensitivity. The theory behind this modification was explained by Kukita *et al.* (1997), who found that TBE buffer with glycerol maintained a relatively low pH (8.1 or less) with improved separation of mutant fragments of up to 800 base pairs.

In addition to the electrophoretic conditions of SSCP, the gel visualisation technique can also influence the sensitivity of the method and detection rates. Radiolabelling of DNA, by end-labelling one of the PCR primers, has traditionally been used to allow visualisation by autoradiography (Nataraj *et al.*, 1999). For simplicity and ease of handling silver-staining is often the method of choice, although it is not always as sensitive as radiolabelling (Budowle *et al.*, 1991). When a high throughput of samples is required, SSCP using fluorescently-labelled primers on automated capillary sequencers can be used (Larsen *et al.*, 1999). Running times on conventional gels can be improved using fluorescent scanners with dyes/stains such as SYBR Gold (Molecular Probes) (Tuma *et al.*, 1999).

### III.1.ii(b). HDA, CSGE, DGGE and DHPLC.

Other electrophoretic gel-based screening techniques, akin to SSCP in their simplicity and dependence upon conformational change, include heteroduplex analysis (HDA) (Nagamine, *et al.*, 1989; White *et al.*, 1992), conformation sensitive gel electrophoresis (CSGE) (Ganguly *et al.*, 1993), and denaturing gradient gel electrophoresis (DGGE) (Fischer & Lerman, 1979). HDA works on the principle that mismatched DNA duplexes, *i.e.* heteroduplexes, have an altered electrophoretic mobility compared to homoduplexes (White *et al.*, 1992). Heteroduplexes form when PCR products from a heterozygote template re-anneal and can be distinguished from homoduplexes by electrophoresis. The conformation changes which result from single base pair alterations (a bubble-type) can be much more difficult to detect by this method than a small deletion or insertion (bulge-type) (Battacharyya & Lilley, 1989). HDA has been used to detect mutations in the unique and duplicated regions of *PKD1* (Table VII.1,

p252). Mutation detection rates can improve when a combination of HDA and SSCP are used, *e.g.* with the neurofibromatosis type 1 gene (*NF1*) (Abernathy, *et al.*, 1997). CSGE is a modification of HDA with 15% formamide and 10% ethylene glycol used to emphasise conformational differences (Ganguly *et al.*, 1993).

DGGE works on the principle of partial melting of double-stranded DNA in polyacrylamide gels (Fischer & Lermann, 1979). In the presence of urea and formamide gradient, DNA which is A:T-rich denatures more readily than G:C-rich regions, thus altering the electrophoretic mobility of the DNA (Fischer & Lermann, 1979). If the sample is heterozygous, the four possible homoduplex and heteroduplex species formed during PCR amplification melt at different rates to produce four bands on the polyacrylamide gel, compared to the one band seen in the wild-type control (Cotton, 1997). If, however, the DNA is homozygous for a particular sequence variation it produces only one band, which is different to that seen in the wild-type homozygous specimen. Mutations in *PKDI* have been detected using DGGE, although the method generally requires significant optimisation and can be difficult to use because of the high GC content of the gene (Perrichot *et al.*, 1999).

Denaturing high performance liquid chromatography (DHPLC) involves the liquid chromatography of PCR products through a column containing alkylated nonporous particles (Hayward-Lester *et al.*, 1996; Liu *et al.*, 1998a). Due to a lower binding capacity, heteroduplexes display reduced column retention time relative to homoduplexes. DHPLC is highly sensitive for fragments between 220 to 445 bp (Arnold *et al.*, 2000). A disadvantage of the technique is that relatively expensive

equipment is required for DHPLC and mobile phase temperatures must be calculated with regard to GC content for all PCR fragments screened (Liu *et al.*, 1998a).

### III.1.ii.(c). Dideoxyfingerprinting

This method is a hybrid of SSCP and Sanger dideoxyfingerprinting (Sarkar *et al.*, 1992). Sanger sequencing reactions, performed in forward and reverse directions using only one dideoxy terminator, are subjected to non-denaturing gel electrophoresis. Mutations present in the sequencing reaction can be detected in the dideoxy component of the reaction because of a gain or loss of a dideoxy termination segment. Since this gain or loss could be masked by the normal allele, the method is further enhanced by the inclusion of a SSCP gel fractionation reaction, which can detect a shift in mobility of at least one of the termination segments that follow the mutation (Blaszyk *et al.*, 1995). The strength of this assay is therefore in the ability to generate a spectrum of fragments for modified SSCP analysis, thus increasing the sensitivity of SSCP.

Improvements to the technique have been made by increasing the length of fragment screened from about 250 bp up to 500 bp, through Bi-directional dideoxy fingerprinting (Bi-ddF). Two simultaneous cycle sequencing termination reactions are performed in the down- and upstream directions, using only one dideoxy terminator and two labelled primers (Haavik *et al.*, 1996; Liu *et al.*, 1996). The most recent update to the protocol, denaturation fingerprinting (dnF<sub>2R</sub>), is suitable for screening regions with a high GC content and utilises DGGE and Bi-ddF with two dideoxy terminators (Liu, *et al.*, 1998b). During the course of the Human Genome Project, relative direct sequencing costs have fallen. Sequencing is now the first choice of many researchers.

### III.1.ii.(d). Mismatch Cleavage Methods

Enzymatic mismatch cleavage (EMC) is another mutation detection screening test which, as its name suggests, utilises the properties of specific enzymes to cleave double-strand heteroduplex DNA at sites of mismatch (Kleff & Kemper, 1988). A T4 Endo VII-based mutation detection kit, the PASSPORT Mutation Scanning Kit (Amersham, Pharmacia Biotech), is now commercially available (Inganäs *et al.*, 2000). In EMC, heteroduplex DNA is formed when an excess of mutant DNA is mixed in the same tube with wild-type DNA, melted, and then allowed to re-anneal. The technique has been shown to detect all 4 possible types of mismatched bases with up to 95% efficiency (Youil *et al.*, 1995). As it can screen in excess of 1 kb (Forrest *et al.*, 1995), it may be preferable to SSCP in terms of the number of reactions required to screen large genes.

A similar technique to EMC is referred to as chemical mismatch cleavage (CMC), which was developed as a modification of the Maxam-Gilbert DNA sequencing method (Cotton *et al.*, 1988). Mismatched thiamines are susceptible to modification by osmium tetroxide and mismatched cytosines by hydroxylamine. The modified bases are then cleaved by piperidine treatment. Unfortunately, osmium tetroxide is a noxious chemical, and for this reason the use of CMC has been somewhat limited (Taylor & Deeble, 1999).

### III.1.ii.(e). DNA Sequencing

In addition to its capacity in characterising sequence variations indicated by other screening techniques, DNA sequencing can also be used as a direct screening technique.

Sequencing can be performed using traditional single-strand manual dideoxy sequencing and radio-labelling (Sanger *et al.*, 1977 & 1980; Gyllensten & Erlich, 1988), or thermal cycle sequencing (Murray, 1989).

The first approach can be undertaken in two ways. The traditional Sanger sequencing method used the Klenow fragment (large subunit of DNA polymerase I) to synthesise a short complementary copy from specifically primed DNA template (Sanger *et al.*, 1977). This was performed with radiolabelled dATP and then transferred into four separate reaction tubes, each containing all four dNTPs and a specific ddNTP. This approach failed to sequence the first 50 or so bases from the primer. Alternatively, a labelling/termination method could be performed (Tabor & Richardson, 1987), where asymmetric PCR products are used as template for the sequencing reactions. Primers are annealed to the template and radiolabelled dNTPs are incorporated into the sequencing reactions using T7 polymerase. For both methods, sequencing reactions are denatured then loaded in adjacent lanes of denaturing polyacrylamide gels. Thermal cycle sequencing is less cumbersome to perform, as fluorescent primers can be used and only one reaction tube per sample is necessary (Murray, 1989).

A dideoxy cycle sequencing reaction is performed when template, primer, dNTPs, ddNTPs and a thermostable polymerase are subjected to repeated rounds of denaturation, annealing and synthesis. Most commercially available sequencing kits for automated sequencing now use fluorescently labelled dideoxynucleotide terminators, relying on fluorescence imaging systems for detection. For example, the Applied Biosystems Incorporated (ABI) Prism Dye terminator chemistry provides a robust non-



isotopic method of sequencing on an automated sequencer (Kwok *et al.*, 1994). DNA sequencing technologies have improved rapidly during the course of the Genome Projects. Advancements and modifications include the invention of high-throughput sequencers, the use of dideoxy-terminators labelled with energy transfer dyes (BigDyes) (ABI), and combination of the annealing and extension steps of sequencing reactions. The new chemistry has increased peak heights and allowed greater sequencing accuracy (Wen, 2001).

### III.1.iii. RNA Screening Techniques.

Another cleavage-based method of mutation screening is non-isotopic RNase cleavage (NIRCA) (Goldrick *et al.*, 1996), developed as a modification to ribonuclease cleavage (Myers *et al.*, 1985). NIRCA works on the principle that duplex RNA targets, up to 1 kb in length, are cleaved by RNase at their site of base pair mismatch (Goldrick *et al.*, 1996). The mismatched substrates are generated by *in vitro* transcription, and PCR or RT-PCR (Reverse Transcription-Polymerase Chain Reaction) amplification is performed using primers with bacteriophage RNA polymerase promoters to subsequently generate large quantities of RNA from both strands (Goldrick *et al.*, 1996). Complementary wild-type and mutant transcripts are hybridised and incubated with RNase. Cleavage products are subjected to electrophoresis and gels are visualised by ethidium bromide staining (Goldrick *et al.*, 1996). Mutations in *PKD1* have been detected using this method (Peral *et al.*, 1997).

### III.1.iv. Protein Screening Techniques.

The protein truncation test (PTT) is usually based on a combination of RT-PCR coupled to *in vitro* transcription and translation, and is used to selectively detect translation-terminating mutations (Roest *et al.*, 1993), it can also be performed on DNA from within large exons. PTT is an appropriate mutation screening method for disorders where the majority of mutations result in premature translation termination. The PTT comprises 4 main steps: (i). RNA is isolated from peripheral blood lymphocytes or other tissue, with the assumption that it is processed in an identical manner to the target tissue; (ii). it is then reverse transcribed to c.DNA; (iii). c.DNA (or DNA if used) is amplified by PCR; (iv). the PCR product is used as template in a 'nested PCR', performed with an inner primer containing a T7-promoter and a eukaryotic translation initiation sequence at the 5' end. This extra sequence allows a single tube *in vitro* transcription and translation of the PCR products. The coupled T7 RNA polymerase transcription and rabbit reticulocyte lysate system (Promega™) results in the incorporation of [<sup>35</sup>S] methionine or other radio-labelled amino acids into the nascent peptides. The translation products are then fractionated by polyacrylamide gel electrophoresis (SDS-PAGE), and visualised by autoradiography (Roest *et al.*, 1993).

Truncated protein products resulting from premature termination are shorter than the normal protein product and migrate more quickly through the gel. Size markers are used to estimate the length of a truncated protein and automated sequencing can then be performed in the region where a mutation is expected. Although this method can only be used when the disease gene is prone to translation-terminating mutations, it is particularly useful if the gene is large and has many exons. PTT has been successfully

used to detect mutations in products varying in length from 971 to 5,131 bp (Roest *et al.*, 1993), including *PKD1* (Roelfsema *et al.*, 1997; Peral *et al.*, 1997). There is a limitation associated with this technique, as it can fail to detect mutations in the presence of nonsense-mediated mRNA decay. This limitation can be reduced by combining the RT-PCR/PTT protocol with the preincubation of cells in cyclohexamide to stabilise the mutant RNA (Bateman *et al.*, 1999).

### III.1.v. Comparison of Mutation Detection Methods.

A recent study by Yuan *et al.* (1999) using the Marfan syndrome *FBNI* gene (encoding a 2,871-aa protein) and the hepatocellular carcinoma P53 gene (encoding a 393-aa protein) compared SSCP/HDA, EMC and direct sequencing in terms of cost and efficiency. When labour and reagents were taken into account SSCP/HDA was the cheapest method with a cost of US\$865.84 per patient, compared to US\$1585.82 per patient for EMC, and US\$1741.02 per patient for direct sequence analysis.

The detection efficiency of the three methods varied. Although SSCP/HDA was the cheapest method of mutation screening, four of the twenty control mutations tested remained undetected (80% detection rate); three of these mutations were contained in PCR fragments greater than 400 bp in size (Yuan *et al.*, 1999). When the direct sequencing method was performed with primers in only one direction two mutations in *FBNI* were missed, this represents a 90% detection rate. Missed mutations were claimed to be due to fluorescent background interference (Yuan *et al.*, 1999). When the samples were re-amplified and directly sequenced in both directions, the mutations were duly detected, the sequencing chemistry used was not described (Yuan *et al.*, 1999). All

20 mutations were detected by EMC. A criticism made by the authors regarding EMC was that considerable time and effort were needed for initial optimisation. Pre-diagnostic optimisation would increase the cost of the analysis when compared to other more widely practised methods (Yuan *et al.*, 1999).

There is a lack of literature on cost and efficiency analyses of different screening methods, although individual studies provide approximate rates of efficiency/detection with various genes. A detection rate of almost 100% has been reported using DGGE after the design of specific primers and the addition of GC clamps (Lerman & Silverstein, 1987; Theophilus *et al.*, 1989; Fodde & Losekoot, 1994). NIRCA has been estimated to have a detection rate of 80-90% (Taylor & Deeble, 1999). A 100% detection rate was achieved using ddF to on the human factor IX gene (Sarkar *et al.*, 1992) and the P53 gene (Blaszyk *et al.*, 1995). It is difficult to estimate the overall efficiency of mutation detection with PTT as it can only be used to detect truncating errors. The PTT detection rate in *NF1* (neurofibromatosis gene) has been reported at 47.1% (40/85) (Fahsold *et al.*, 2000). As 80% of mutations in the *NF1* gene are estimated to be truncating, if this is considered, the relative detection rate of PTT for truncating mutations increases to 58.8% (40/68) (Fahsold *et al.*, 2000).

It can be assumed that the choice of screening technique for any particular mutation detection study must be the one that is expected to achieve a high detection rate and is available within a given budget. Since no screening technique can offer 100% accuracy, and sequencing costs are becoming lower, many researchers are choosing direct sequencing rather than preliminary screening techniques.

### III.1.vi. Amplification of *PKD1* without Homologous Gene (HG) contamination

Methods to prevent the co-amplification of the HGs when screening for mutations in the duplicated region of *PKD1* are summarised in Table III.2. They include: (1) “anchored” primers in the unique region of the gene for long-range amplification into the duplicated area (Watnick *et al.*, 1997; Peral *et al.*, 1997; Koptides *et al.*, 1998; Thongnoppakhun *et al.*, 1999); (2) the PTT method (Roelfsema *et al.*, 1997); or (3) design of *PKD1* specific primers (Watnick *et al.*, 1999; Thomas *et al.*, 1999).

Amplifications from primers anchored in the unique sequences of *PKD1* into the duplicated region of the gene have been successful from exon 34 to primers located upstream of exon 23. It was not considered possible to perform long-range amplification of genomic DNA from exon 34 further 5' of exon 23 as IVS22 contains the longest polypyrimidine tract (approx. 2.5 kb) found in any human gene (Van Raay *et al.*, 1996). Most researchers have found it impossible to amplify across the tract to include exons upstream of exon 23.

The design of *PKD1*-specific primers for use in the duplicated region has been difficult due to a lack of sequence information on the HGs. Although some sequence information has been produced, the Human Genome Project has not completed characterisation of the entire chromosome 16 region where the HGs lie. To solve this problem, sequences in the 5' region of *PKD1* unique to that gene were sought, so that appropriate primers could be designed. It has been shown that the HG identity with *PKD1* does not span the entire 5' region (Watnick *et al.*, 1997). BAC (bacterial

artificial chromosome) libraries can be screened by PCR to locate clones (or “pools” of clones) containing DNA inserts that amplify with the primers of interest.

BACs can be used for cloning large inserts of DNA ranging from 100 to 300 kb (Monaco & Larin, 1994). Clones consist of a bacterial vector with a stretch of foreign insert DNA (Shizuya *et al.*, 1992). There is little chimaerism associated with BACs, so insert DNA is expected to originate from a single source of continuous DNA (Monaco & Larin, 1994). This implies that when the clones in a BAC library are screened for *PKDI* sequences, some may correspond to the *PKDI* gene and others to one or more of the HGs. No single BAC clone could contain both *PKDI* and the HGs, as these sequences are physically too far apart on chromosome 16.

BAC libraries are arranged in 96-well microtitre plates and correspond to particular chromosomes. The super-pool plates have DNA from eight individual clones pooled together per well. A detectable product indicates that one or more of the clones in the super-pool plate contain the specific insert of interest, so corresponding pool plates can then be screened to determine which clones are of interest. Clones are then obtained as glycerol stocks, cultured, and the DNA is extracted. This DNA originating from *PKDI* or the HGs (but not both) is sequenced, and variations in DNA sequence can be used to design *PKDI* specific primers.

### III.1.vii. Methodology specific to ADPKD

Most studies of mutations in the *PKDI* gene have screened only the 3' unique region (Table III.1). The groups that report mutations in the duplicated region (Table III.2)

have used slightly different strategies to prevent co-amplification of the HGs and detect mutations in *PKD1*.

**Table III.1.** Mutation Screening of DNA in Exons 34 to 46 of the *PKD1* Gene.

Screening Method	Reference	Mutations detected/ ADPKD Individuals tested
SSCP	Peral <i>et al.</i> , (1996b)	2/45-6/150
	Peral <i>et al.</i> , (1996a)	1 family only
	Neophytou <i>et al.</i> , (1996)	1/14
	Torra <i>et al.</i> , (1997)	1 family only
	Daniells <i>et al.</i> , (1998b)	5/147
	Badenas <i>et al.</i> , (1999)	7/175
	Afzal <i>et al.</i> , (1999)	4/90
	Kim <i>et al.</i> , (2000b)	6/91
		<b>Total = 31/669</b>
HDA	Peral <i>et al.</i> , (1995)	2/130
	Turco <i>et al.</i> , (1995)	1/20
	Rossetti <i>et al.</i> , (1996)	1/67
	Aguiari <i>et al.</i> , (2000)	3/40
		<b>Total = 8/257</b>
HDA & SSCP	Turco <i>et al.</i> , (1997)	3/60
		<b>Total = 3/60</b>
DHPLC & SSCP	Mizoguchi <i>et al.</i> , (2001)	4/176
		<b>Total = 4/176</b>
DGGE	Perrichot <i>et al.</i> , (1999)	8/146
	Perrichot <i>et al.</i> , (2000a)	additional 3/146
		<b>Total = 11/146</b>
PTT & Direct Sequencing	Rossetti <i>et al.</i> , (2001)	6/131
		<b>Total = 6/131</b>
RT-PCR & PTT	Roelfsema <i>et al.</i> , (1997)	1/20
		<b>Total = 1/20</b>
Grand Total		<b>Grand Total = 64/1459 (4.4%)</b>

**Table III.2.** Mutations Detected in the Duplicated Region of the *PKD1* gene.

Screening Method ** (& Template)	Reference	Mutations Detected/Tested Individuals	Exons Screened in Duplicated PKD1 Region
SSCP (DNA)	Koptides <i>et al.</i> , (1998)	1 patient only	Ex23-34
	Watnick <i>et al.</i> , (1999)	8/35	Ex 11-21
	Afzal <i>et al.</i> , (2000)	3/90	Ex 23-34
	Koptides <i>et al.</i> , (2000)	1 patient only	Ex 23-34
	Bouba <i>et al.</i> , (2001)	8/53	Ex 16-34
	Phakdeekitcharoen <i>et al.</i> , (2000 & 2001)	19/54	Ex 1-34
		<b>Total = 40/234</b>	
HDA (DNA)	Watnick <i>et al.</i> , (1997)	8/80	Ex 23-34
		<b>Total = 8/80</b>	
DGGE (DNA)	Perrichot <i>et al.</i> , (2000b)	8/146	Ex 15-21 & 23-34
		<b>Total = 8/146</b>	
PTT ( <i>in vitro</i> translated)	Roelfsema <i>et al.</i> , (1997)	6/135	Ex 15 & 23
		<b>Total = 6/135</b>	
PTT & NIRCA (RT-PCR)	Peral <i>et al.</i> , (1997)	6/95	Ex 22-46
		<b>Total = 6/95</b>	
PTT & Direct Sequencing	Rossetti <i>et al.</i> , (2001)	51/131	Ex 1-34
		<b>Total = 51/131</b>	
Direct Sequencing			
DNA	Thomas <i>et al.</i> , (1999)	7/24,	Ex2-15
RNA	Thongnoppakhun <i>et al.</i> , (2000)	1 patient only	Full length RNA
		<b>Total = 8/25</b>	
Grand Total		<b>Grand Total = 127/846 (15.0%)</b>	

\*\*Screening was usually performed on dilute XL-PCR *PKD1*-specific template, from primers anchored in the unique region of the gene or by using sequences in the duplicated region specific to *PKD1*.



The European PKD Consortium (1994) first reported mutations in the duplicated region of *PKD1*, although they did not specifically screen for 5' mutations. Rather, they were able to confirm that the PBP (Polycystic Break Point) gene was the *PKD1* gene following the discovery of a total deletion in a Portuguese family. Three additional aberrant transcripts were then detected in *PKD1*, namely a 2 kb deletion between exons 31-34, a 5.5 kb deletion from exon 32 to the 3'UTR, and a splicing mutation in IVS 44.

The first study to specifically screen for mutations in the 5' region used PTT to detect six truncating mutations in *PKD1* exons 15 and 23 (Roelfsema *et al.*, 1997). On the assumption that the HGs were transcribed but not translated (European PKD Consortium, 1994), only the *PKD1* protein product was screened by PTT and the HGs were not translated (Roelfsema *et al.*, 1997). The approach was somewhat limited because amplification required exonic primers, and only the larger exons of *PKD1*, e.g. 15 and 23, could be screened using genomic DNA as template. In addition, missense mutations, which may account for approximately 30% of mutations in the duplicated region of *PKD* (see Chapter II), would not be detected by PTT (Roelfsema *et al.*, 1997). Peral *et al.* (1997) also used a combination of PTT and NIRCA to detect six mutations in cDNA templates amplified by RT-PCR. The 2.4 kb template was amplified with a *PKD1*-specific primer in the unique region of the gene (exon 34) and a forward primer in the duplicated region (exon 22).

Most groups have used genomic DNA as template for mutation screening. Anchored *PKD1*-specific primers in the unique 3' region were used with primers which extend

into the repeat region of the gene for long range amplification (XL-PCR), followed by various screening methods (Watnick *et al.*, 1997; Peral *et al.*, 1997; Koptides *et al.*, 1998 & 2000; Perrichot *et al.*, 2000b; Afzal *et al.*, 2000; Phakdeekitcharoen *et al.*, 2000 & 2001; Bouba *et al.*, 2001). The *PKD1*-specific XL-PCR amplified product was diluted ( $> 1:10^4$ ) prior to re-amplification, with nested primers to overcome the problem of any residual genomic contamination by HGs from the initial PCR amplification (Watnick *et al.*, 1997).

The first two groups to detect *PKD1* mutations between exons 2 to 5 and 11 to 21 of the *PKD1* gene used XL-PCR from *PKD1*-specific primer sites in IVS1 and exon 15 (Thomas *et al.*, 1999; Watnick *et al.*, 1999). Long-range PCR products were then diluted and used as template for the amplification of individual exons. Specific *PKD1* primers were designed by comparing the known *PKD1* sequence to a BAC clone (GenBank AC002039) that contained 2 copies of the HGs (Loftus *et al.*, 1999).

More recently two groups (Rossetti *et al.*, 2001; Phakdeekitcharoen *et al.*, 2001), reported screening the entire duplicated region of the *PKD1* gene. In the Rossetti study, 51 mutations were detected between exons 1 to 34 from 131 individuals tested with a combination of the PTT and direct sequence analysis. Phakdeekitcharoen *et al.* (2001) compared the *PKD1* sequence to the published sequence of HGs 1 & 2 (Loftus *et al.*, 1999), in order to design *PKD1*-specific primers sites. From these results they were able to amplify exons 1 to 34 as eight XL-PCR fragments. Exon 22 was amplified as a separate exon to avoid amplification across its surrounding polypyrimidine tracts. XL-

PCR fragments were then diluted ( $> 1:10^4$ ) and used as template in nested-PCR followed by SSCP analysis.

One group has reported the first long-range RT-PCR of the entire coding sequence of the *PKD1* gene (Thongnoppakhun *et al.*, 1999), with the XL-PCR cDNA template diluted 1000-fold prior to nested PCR to avoid HG contamination. This led to the discovery of a splice-acceptor site mutation, in IVS 13 of *PKD1*, by direct manual sequencing (Thongnoppakhun *et al.*, 2000).

Afzal & Jeffery (2001), recently reported the XL-PCR amplification of a 13.5 kb *PKD1* fragment between exon 34 and IVS21. An amplification of this type from genomic DNA was previously believed to be impossible, due to difficulties involved in extending amplification over the 2.5 kb polypyrimidine tract in IVS 22 (Van Raay *et al.*, 1996). The present study was not able to replicate the results for this amplification. If this XL-PCR amplification could be performed routinely, it would allow full screening of the *PKD1* gene between exons 22 and 34 without HG interference.

Alternatively, other strategies could be developed which permit the amplification of *PKD1* without co-amplification of the homologous genes. Either way, it would mean that more mutations could be detected throughout the gene and a significant genotype-phenotype correlation study would be possible. Outcomes of this type of research would lead to a greater understanding of the mechanisms of mutation in ADPKD, the function of polycystin-1, its possible pathways and interactions, the disease pathogenesis, and the wide-ranging phenotypic variability found in ADPKD.

### **III. Methodology (Part II)**

#### **III.2-7. Subjects and Methods**

##### **III.2. Additional Protocols**

Additional methods used in this project may be found in Appendix A. Details of protocols relevant to the thesis, accompanied by the various solutions used, are presented in that section.

##### **III.3 Subjects**

###### **III.3.i. Sample Collection and Preparation Prior to Analysis**

Researchers other than this candidate undertook the preliminary set-up of the study, phenotypic data and clinical samples were collected, genomic DNA from peripheral blood was extracted, and linkage analysis studies were performed. Results of this work were made available to this candidate. Except where specifically indicated, the author conducted all further practical work and statistical analyses.

Genomic DNA or blood spots from patients tested as part of this project were obtained from three collaborative studies. Australian, Bulgarian and Polish families were recruited for a multi-centre study investigating the genetic and environmental factors involved in ADPKD. Families were ascertained through the patient registry at the Department of Nephrology, Royal Perth Hospital (RPH), Perth, Western Australia; the haemodialysis centres at the Departments of Nephrology in Plovdiv and Pleven, Bulgaria; and the Department of Nephrology, Jagiellonian University, Krakow, Poland.

The diagnosis of ADPKD followed standard ultrasonographic criteria first described by Bear *et al.* (1984), then further refined by Ravine *et al.* (1994). The same diagnostic protocol was used in all three countries. When patients gave informed consent to participate in the study, they were invited for interview at the various centres and familial and clinical histories were established. Information on the nature and pattern of inheritance of the disease was provided to patients with preliminary genetic counselling, additional counselling was available if requested. A pedigree was constructed for each patient and questionnaires on genetic, clinical, dietary and lifestyle factors were administered. The results of these questionnaires, in combination with clinical tests, provided the basis of the phenotypic information made available to this candidate. All index patients were requested to invite family members to participate in the study, and most families responded positively with affected and unaffected (or unknown status) individuals agreeing to join the study. The information gathered from patients included general questions, such as date and place of birth, ethnic origin, number and disease status of family members. This was followed by specific clinical questions on diagnosis and reason for examination, first symptom (and what that symptom was), onset of hypertension, CRF, and ESRF, details of kidney transplantation, occurrence of extra-renal cysts and other extra-renal symptoms. The ages at which these events occurred were also noted.

To ensure confidentiality, a numbering system, rather than family name was used to identify all participant data and biological samples in the study. Clinical data were also obtained from medical records with patients consent. These data included ultrasound

findings, blood pressure values, and biochemical analyses including serum creatinine levels, plus the type and severity of symptoms at diagnosis and at subsequent visits.

Linkage analysis was performed when possible by Ms Tina Schiavello (Centre for Human Genetics, Edith Cowan University, Perth, Australia), Prof Joachim Hallmayer (University of Western Australia, Perth, WA) and Dr Nadja Bogdanova (Institute of Human Genetics, Westfälische Wilhelms-Universität (WWU), Münster, Germany), results of their work were made available to this candidate. Using this information, it was possible to screen those families who had ADPKD most probably linked to the *PKD1* gene on chromosome 16, thus increasing the chance of mutation detection. Observed minimum and maximum LOD scores in those “positively” *PKD1*-linked families ranged from 0.3-1.8 and all families had a probability of linkage to PKD1 of 0.9 or greater using the methods described by Hateboer *et al.* (1999b).

The study populations consisted of 26 families from Bulgaria, 19 from Australia and 7 from Poland. In addition to the 19 Australian families, 31 Australian ADPKD index patients of unknown linkage status were included for direct sequence analysis of exons 44-46 of the *PKD1* gene.

Genomic DNA samples were either extracted from peripheral whole blood or blood spots, or made available as a blood spot on filter paper. The DNA extractions of whole blood were performed by Dr Bogdanova and Ms Schiavello using a salt precipitation method (Miller *et al.*, 1988), and DNA from blood spots was extracted using a phenol/chloroform method (Sambrook *et al.*, 1989).

The extracted genomic DNA was measured by UV-light spectrophotometry at 260nm to determine its concentration and to assess DNA purity. Stock solutions were stored at -70°C, and working DNA solutions were prepared at concentrations of 50ng/μl and 10ng/μl and stored at 4°C.

It was possible to directly PCR amplify fragments of 200-400 bp in length of genomic DNA from blood spot filter papers without prior extraction. A blood spot (approximately 3 mm in diameter) was cut from the filter paper and placed into the PCR reaction tube. An extra 5μl of ddH<sub>2</sub>O was added to the 25μl total reaction volume to replace the 5μl volume of DNA typically used. This type of PCR was only performed on some family members who were unable to provide whole blood samples for analysis. All index patients in the study donated whole blood for DNA extraction.

#### III.4. Mutation Detection of *PKD1* - 3' Unique Region

The region spanning exons 34 to the 3'-UTR of *PKD1* is unique to the gene, therefore co-amplification of homologous genes was not a problem. However, *PKD1* is GC-rich and some exons required a high degree of optimisation prior to successful amplification, after which mutation screening was performed using SSCP analysis. Any variations detected in SSCP banding patterns were further investigated by direct sequencing. Where possible, suspected disease-causing sequence variants were tested for familial segregation.

Mutations not resulting in premature truncation, *i.e.* missense variants and in-frame deletions and insertions, were subjected to secondary protein structure programs such as

the PHDsec algorithm (Rost *et al.*, 1994) or nnPredict (BCM HTTP address) in order to predict their likely effect on protein structure. Evolutionary conservation of DNA sequence variations was assessed with BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) or Pfam (<http://www.sanger.ac.uk/Software/Pfam/>). Phenotypic data for recurrent mutations and mutations in identical protein domains were compared between and within families.

#### III.4.i. Polymerase Chain Reaction

PCR was used to amplify genomic DNA for the individual exons in the unique region of the *PKD1* gene. The basic amplification reagents are detailed below (Table III.3) and specific primers and conditions to each amplification reaction are listed in Table III.5.

**Table III.3** PCR Amplification Reagents.

PCR Reaction Mix (x1 volume)		
Reagent (conc.)	Volume	Final Conc.
Template genomic DNA (10ng/μl)	5.0μl	2ng/μl
10X Buffer	2.5μl	x1
MgCl <sub>2</sub> (50mM)	0.8μl	1.6 mM
1.25mM each dNTPs	4.0μl	200μM each/rxn
Forward Primer (60ng/ul)	0.5μl	1.2ng/μl
Reverse Primer (60ng/ul)	0.5μl	1.2ng/μl
<i>Taq</i> Polymerase (5.5U/μl)*	0.2μl	0.044U/μl
ddH <sub>2</sub> O up to 25μl	11.5μl	
<b>Total Volume</b>	<b>25μl</b>	

(\*Unless stated standard *Taq* DNA Polymerase and accompanying 10X Reaction Buffer for PCR were purchased from Life Technologies-Gibco BRL or QIAGEN).



Examples of the thermocycling conditions for PCR are given below (Table III.4), exon specific conditions are listed in Table III.5. Unless otherwise stated, all PCR reactions have three cycling steps (denaturation, annealing and extension). A two-step PCR indicates the same temperature for both the annealing and extension steps.

**Table III.4. PCR Cycling Steps**

<b>30 cycles at 65°C (3-step)</b>		<b>35 cycles at 68°C (2-step)</b>	
95°C for 3 mins	x1	95°C for 3 mins	x1
65°C for 1 mins	}	68°C for 2 mins	}
72°C for 2 mins	}x30	95°C for 0.5 mins	}x35
95°C for 0.5 mins	}		
65°C for 1 min	x1	68°C for 7 mins	x1
72°C for 7 mins	x1		

Despite repeated attempts to PCR amplify exon 42, a specific product for this region could not be generated. The various primers used are listed in Table III.6. In addition to the routine PCR reaction mix, additives such as DMSO, BSA plus gelatine, betaine, Q-solution (QIAGEN) and Perfect-Match PCR enhancer (Stratagene), and altered cycling conditions with annealing temperatures ranging from 55°C to 68°C were also evaluated. A series of buffer solutions (A-O see Appendix A) with differing MgCl<sub>2</sub> concentrations were also unsuccessful in generating the expected amplification product. Attempts to amplify a product of 821bp across exon 42 (from exon 41 to exon 43) was also unsuccessful, even after employing nested PCR.

Table III.5. PCR of *PKDI* Unique region

Exon	Primers	Annealing Temperature & Special Conditions	Size (bp) of Product	Genomic Position (L39891)
34	F 5' gtgagctggggtgagaggag 3' R 5' acggctgcttggcctgagtc 3'	65°C (X 30) (BSA & Gelatine)	240	44276-295 44496-515
35	F 5' ctgcaactgcctcctggagg 3' R 5' agctcacaggaggaggtag 3'	68°C (X 35) 2-step	219	47330-349 47530-549
36	F 5' ctgtgagctgcctctcacag 3' R 5' cccgtgatggaggcctgtag 3'	63°C (X 30)	286	47541-560 47807-826
37	F 5' ggtaggtacaggcctccat 3' R 5' ggagacaagagacggagggtg 3'	68°C (X 35) 2-step	299	47801-820 48080-099
38	F 5' caaagccctgctgtcactgt 3' R 5' cctagggtctggctggacta 3'	64°C (X 30)	272	48439-458 48691-710
39	F 5' gtctctggtgccgctcact 3' R 5' agagctccgctaaaggctgc 3'	68°C (X 35) 2-step	259	48929-948 49167-187
40	F 5' gagctggccagcaggaaaca 3' R 5' ctccctgtccttggcgtaga 3'	68°C (X 35) 2-step	264	49330-349 49573-593
41	F 5' cgtctacgccaaggacaagg 3' R 5' tggctgaggggctgtggaag 3'	68°C (X 35) 2-step	355	49571-590 49906-925
42	Failed to amplify despite repeated attempts			
43	F 5' cagcgtccctcccgcctcctga 3' R 5' gcttgagccctgggggtgtgcgc 3'	60°C (X 40) 4% DMSO & 10% 5M Betaine	383	50381-403 50741-763
44	F 5' caccacagggtgcaagcag 3' R 5' acgctccagaggagtcac 3' (3A3 C2 EPKDC, 1994)*	68°C (X 35) 2-step	292	50746-765 51018-037
45	F 5' cgtcttagctcagctcagct 3' R 5' gtgtccctctccccccact 3'	65°C (X 30)	396	50946-965 51322-341
46 (a)	F 5' gtgcagccggactgactgag 3' R 5' gctgtccagctggtagacg 3'	68°C (X 35) 2-step with TMAC	318	51361-380 51659-678
46 (b)	F 5' gaccgactcaaccaggccac 3' R 5' cgtgcagccattctgctg 3' (PTT4, Peral 1997)**	66°C (X 35)	336	51633-652 51950-968

\*Previous reference for exon 44 primer (3A3 C2) European Polycystic Kidney Disease Consortium, (1994).

\*\* Previous reference for exon 46 primer (PTT-4 reverse) Peral *et al.* (1997).

**Table III.6. Exon 42 Amplification Primers**

Primer Description	Primer Sequence	Genomic Position	Size of Product
Initial Exon 42	Forward 5' ettccacagccctgagcca 3' Reverse 5' ctgcgaggggtgagacgtg 3'	49906-925 50244-263	358 bp
Exon 41- Exon 43	Forward 5' agctgcacaactggctggac 3' (Peral <i>et al.</i> , 1996b) Reverse 5' cgcacgaaacgggtccac 3' (Peral <i>et al.</i> , 1996b)	49789-808 50592-609	821 bp
Nested Exon 42	Forward 5' ctgagccacgctgcactgc 3' Reverse 5' ctgcgaggggtgagacgtg 3'	49918-937 50244-263	346 bp

#### III.4.ii. Agarose Gel Electrophoresis

To confirm successful PCR amplification and the absence of primer-dimer contamination, 5µl aliquots of all PCR products in Ficoll loading buffer were loaded on a 2% agarose gel containing Ethidium Bromide (0.5µg/ml). Electrophoresis was performed in wide Mini-Sub™ cells (BIO-RAD), with gels (15 x 10 x 0.7cm; 2 x 20 wells) run in 1 x TAE buffer at a constant 80V for 40 to 60 mins. A suitable size marker (*eg.* 100bp DNA ladder, Gibco BRL) was also loaded on to the gel to ensure that the amplified products were of the expected size. After electrophoresis, the gels were visualised on a bench-top UV-transilluminator (Mighty Bright UV-transilluminator, Hoefer Instruments), and a polaroid photograph or digital image was taken as a permanent record.

#### III.4.iii. Single Strand Conformation Polymorphism Analysis

Single strand conformation polymorphism analysis (SSCP) was chosen as the method for mutation screening of the unique region of *PKD1*. After testing different cross-linking ratios, it was determined, that a 12% Polyacrylamide gel (acrylamide cross-linked with bis-acrylamide at a ratio of 29:1) was the most suitable for SSCP as previously reported (Savov *et al.*, 1992). Gels were poured between vertical BIO-RAD

glass plates (using the BIO-RAD gel pouring stand) to produce a gel size of 16cm x 16cm x 0.5cm, with a 16-well comb. When polymerised, gels were assembled on the PROTEAN™ II xi Vertical Electrophoresis Cell apparatus (BIO-RAD). The running buffers were 1XTAE or 1XTBE (with 5% glycerol to prevent an increase in the pH of TBE over 7.7) (Kukita *et al.*, 1997).

PCR products (2µl aliquots) were prepared for SSCP analysis by the addition of equal volumes of formamide loading buffer and made up to a final volume of 6µl with ddH<sub>2</sub>O. Samples were denatured in a water bath at 95°C for two minutes and then placed immediately on ice prior to loading. Gels were electrophoresed at various temperatures between 4°C and 25°C. It was determined that gels run at 25°C gave the best resolution. The gels were run in the Protean™ apparatus for varying lengths of time at different constant voltages (see Table III.5 for exon-specific conditions) using a BIO-RAD power supply (PowerPac 1000). The voltage and time were considered to be optimal when a clear gel image was obtained by silver staining, and a good separation of bands had been achieved.

The use of positive *PKD1* mutation/polymorphism controls for SSCP optimisation was not possible, as prior to analysis we had no known sequence variants for any of the fragments screened. All fragments were electrophoresed in both 1XTAE and 1XTBE (+ 5% glycerol) as the sensitivity of these buffers occasionally seemed to vary between PCR fragments. In most exons, both buffers were equally useful (as listed in Table III.7), however screening results of the 263 bp fragment containing exon 40 indicated that different conformational changes were detected when the two buffers were



compared. Thus choice of running buffer should be considered when establishing SSCP running conditions for any fragment.

The efficiency of the initial SSCP optimisation was later tested by comparing SSCP results with automated direct sequence analysis of the 3' unique region (see methodology III.4.viii). It was found that all disease-segregating mutations in the re-screened sub-group detected by direct sequence analysis had been indicated in the initial SSCP screen (See Chapter V). However, a neutral polymorphism in a relatively large PCR fragment (396bp) was not detected by SSCP (See Chapter VI).

**Table III.7. Exon-Specific SSCP Conditions**

Exon	12% PAG and Preferred Buffer	Running Time	Running Voltage
34	1 X TBE + 5% glycerol	8 hours	300 V
35	1 X TAE or 1 X TBE +5% glycerol	8 hours	300 V
36	1 X TAE or 1 X TBE +5% glycerol	14 hours	150 V
37	1 X TAE or 1 X TBE +5% glycerol	8 hours	300 V
38	1 X TAE or 1 X TBE +5% glycerol	6 hours	300 V
39	1 X TBE + 5% glycerol	8 hours	300 V
40	1 X TAE & 1 X TBE +5% glycerol	7 hours	300 V
41	1 X TBE + 5% glycerol	14 hours	150 V
43	1 X TAE or 1 X TBE +5% glycerol	15 hours	200 V
44	1 X TAE or 1 X TBE +5% glycerol	7 hours	300 V
45	1 X TAE or 1 X TBE +5% glycerol	15 hours	200 V
46 (a)*	1 X TAE + 0.1% SDS	7 hours	300 V
46 (b)*	1 X TAE or 1 X TBE +5% glycerol	14 hours	200 V

Exon 46 is relatively long (~1.5 kb) and was only amplified as 2 fragments that spanned the start of the exon to the start of the 3'-UTR (a & b).

III.4.iv. Silver Staining

After electrophoresis the SSCP pattern was visualised by a silver-staining procedure (modified from Budowle *et al.*, 1991) as detailed below.

### Silver staining protocol

1. The gel was placed in 10% Ethanol (EtOH) and gently agitated for 5 minutes before the EtOH was removed.
2. The gel was then covered in 1% nitric acid oxidising solution and agitated for 3 minutes before rinsing in distilled water.
3. The gel was placed in 0.1% silver nitrate and agitated for 20 minutes.  $\text{AgNO}_3(\text{aq})$  was then removed and the gel rinsed 3 times in distilled water for 2 minutes.
4. A small amount of developer was added (sodium carbonate anhydrous 0.56M and 0.04524% formaldehyde) until it just covered the gel. The gel was agitated and when the developer began to turn brown it was replaced with fresh developer and further agitated until an adequate signal appeared.
5. When the staining was sufficiently intense to see the marker and SSCP bands, the gel was quickly rinsed in distilled water, drained and immediately covered in 10% glacial acetic acid for at least 10 minutes to stop further development of the signal and fix the silver stain.
6. Acetic acid was drained and the gel covered in distilled water.

Gels could be stored for one to two days in distilled water containing 5% glycerol, or dried overnight in cellulose film using the Gel Drying Kit™ from Promega.

### III.4.v. Single Strand DNA Amplification

When conformation changes were observed by SSCP, sequencing was performed to characterise the precise variation from the published sequence (GenBank Acc. L39891). Manual isotopic sequencing was chosen as the sequencing method in the first instance,

as heterozygotes are easily characterised for autosomal dominant diseases such as ADPKD and, at the time of analysis, it was a readily available method. In later experiments when SSCP efficiency was evaluated, automated sequencing was performed.

Prior to manual isotopic sequencing, a single-stranded DNA (ssDNA) template was prepared from double-stranded PCR product template by asymmetric amplification (Gyllenstein & Erlich, 1988). The remaining PCR product, initially amplified for SSCP analysis, was purified to remove any excess primer or non-specific product prior to this procedure. QIAquick PCR purification kit (QIAGEN) was used as per the manufacturer's recommendations and the clean dsDNA was eluted in 30µl of pure water (Appendix A).

The primers used for this ssDNA amplification were the same as those already listed in Table III.5. The cycling conditions for each exon were retained for both the forward and reverse reactions, however the number of cycles was increased to 40. The detailed protocol for ssDNA amplification (Table III.8) is an example of the forward reaction; reverse product was obtained with an excess of reverse primer.

After amplification the product was concentrated and excess oligonucleotides were removed using Centricon™ centrifugal filter columns (Millipore, Centricon-30) as per manufacturer's recommendations (See Appendix A).

**Table III.8.** Asymmetric PCR Amplification Reagents.

Asymmetric PCR Reaction Mix (x1 volume)

Reagent (conc.)	Volume	Final Conc.
Template DNA	1.0 $\mu$ l	
10 x Buffer (with MgCl <sub>2</sub> )	10.0 $\mu$ l	x1 (1.5mM)
1.25mM dNTPs	16.0 $\mu$ l	200 $\mu$ M each
Forward Primer (60ng/ $\mu$ l)	10.0 $\mu$ l	600ng
Reverse Primer (6ng/ $\mu$ l)	1.0 $\mu$ l	6ng
<i>Taq</i> Polymerase (5.5U/ $\mu$ l)*	0.5 $\mu$ l	2.75U
ddH <sub>2</sub> O up to 100 $\mu$ l	61.5 $\mu$ l	
<b>Total Volume</b>	<b>100<math>\mu</math>l</b>	

(The 10 X Buffer and *Taq* polymerase were from Gibco BRL).

#### III.4.vi. Isotopic Sequencing Reaction

When SSCP analysis indicated a conformational variation, fragments were directly sequenced in both forward and reverse directions using the oligonucleotide internal primers listed in Table.III.10.

Samples were radiolabelled with [ $\alpha$ -<sup>35</sup>S]-dATP (Amersham) for sequencing. This isotope was chosen in preference to [ $\alpha$ -<sup>32</sup>P]-dATP, as its low energy  $\beta$  emissions result in relatively sharper autoradiographic bands. Although increased exposure times are needed for autoradiography with <sup>35</sup>S than <sup>32</sup>P, its longer half-life means that sequencing products labelled with <sup>35</sup>S can be stored for longer periods prior to electrophoresis, compared to those labelled with <sup>32</sup>P. Users of <sup>35</sup>S also are exposed to a lower dose of radiation than with <sup>32</sup>P (Ausubel & Albright, Current Protocols in Molecular Biology Chp 7 suppl 26, 1994). The protocol for this method of sequencing is outlined in Table III.9.



**Table III.9. Isotopic Labelling Reaction.**

In microcentrifuge tube mix the following:

ssDNA	10 $\mu$ l	
Annealing buffer	2 $\mu$ l	
Sequencing primer (6ng/ $\mu$ l)	2 $\mu$ l	
<b>Total volume</b>	<b>14<math>\mu</math>l</b>	[Incubate at 80°C for 5 minutes and allow to cool to room temperature]

A “master-mix” was prepared in a designated radioisotope laboratory in compliance with standard safety procedures. Use of a “master-mix” reduces the handling of radioisotope and minimises error from small volume pipetting. The volume of each reagent in the master-mix was equal to the volume for one sample multiplied by (n + 1) where n was the number of samples.

Preparation of an Isotope Master Mix for Labelling  
(X 1 volume)

Labelling Mix* (1:5)	3.0 $\mu$ l
ddH <sub>2</sub> O	0.5 $\mu$ l
[ $\alpha$ - <sup>35</sup> S]-dATP (10 $\mu$ Ci/ $\mu$ l)	<u>0.5<math>\mu</math>l</u>
<b>Total volume</b>	<b><u>4.0<math>\mu</math>l</u></b>

Add 4.0 $\mu$ l of isotope mix to the template mix and then add T7 DNA polymerase

Template mix	14 $\mu$ l
Isotope mix	4.0 $\mu$ l
T7 Polymerase (1U/ $\mu$ l)	2.0 $\mu$ l
<b>Total volume</b>	<b>20.0<math>\mu</math>l</b>

Stand at room temperature for 5 minutes prior to sequencing reaction.

\* Labelling Mix\* (1:5) (Epicentre Technologies, Madison, WI, USA)

**Table III.10.** Sequencing of *PKDI* Unique region

Exon	Primers (Forward and Reverse)	Genomic Position (L39891)
34	5' caggctgctgagcaggtccg 3' R (reverse) (use with forward primed ssPCR product) 5' gctctgaagctcaccttgc 3' F (forward) (use with reverse primed ssPCR product)	44428-44447 44299-44318
35	5' gctacgcaagcacacctgtc 3' R 5' cgggatgaattcacagccta 3' F	47498-47517 47351-47370
36	5' ccctggcagccccctcacct 3' R 5' ctgtctctgcttccccagga 3' F	47563-47582 47782-47801
38	5' gtgcagctctgcttgatgg 3' R 5' accagagcctcctgggtgac 3' F	48595-48614 48496-48515
39	5' cagccctgccagctca 3' R 5' gtgtgactgatgctgtggca 3' F	49116-49132 48981-48990
40	5' cgtccccgagccattgtg 3' ** (FFR) R 5' gggctctgtgtgttcagca 3' F	49501-49518 49388-49407
41	5' cctgtgtccagccagttgtg 3' ** (NNR) R 5' gagtagttctccaggagtgc 3' F 5' cagccaccctctccggca 3' F	49794-49814 49591-49610 49667-49686
44	5' ccgcgcagtcacctacca 3' R 5' cagattctcgtccgcaggct 3' F	50918-50935 50767-50787
45	5' gggccgtaccaccctcctt 3' R or 5' aggaacaactccacctctcgtag 3' ** (HHR) R 5' cagctcgtgtcttctgtgt 3' F or 5' ggctggggctgttattctc 3' ** (MMF) F or 5' tctacctgtgtcctgccag 3' ** (HHF) F	51303-51321 51242-51265 51000-51019 51163-51182 51078-51098
46 (a)	5' gggagggtcaggctcacac 3' ** (JJR) R 5' gtccgccacaaagtccgct 3' F or 5' tccgtttgaaggatggag 3' ** (JJF) F	51575-51594 51398-51417 51412-51431
46 (b)	5' ggactaagtgtctgtggggt 3' R or 5' caaggcggctgggcagtg 3' ** (KKR)	51850-51896 51760-51777

\*\* Previous reference (with primer name) from Peral *et al.* (1996b).

### III.4.vii. Direct Sequence Analysis (Isotopic)

#### III.4.vii.(a). Sequencing Reaction

After radiolabelling of the ssPCR product, sequencing reactions were performed with termination mixes prepared as directed by the manufacturer (Boehringer Mannheim GmbH) for each of the 4 nucleotide-mixes (Table III.11).

**Table III.11. Termination Mixes for Sequencing**

Reagent	A mix	C mix	G mix	T mix
dATP	150μmol/L	150μmol/L	150μmol/L	150μmol/L
dCTP	150μmol/L	150μmol/L	150μmol/L	150μmol/L
(7-deaza) dGTP	150μmol/L	150μmol/L	150μmol/L	150μmol/L
dTTP	150μmol/L	150μmol/L	150μmol/L	150μmol/L
ddATP	15μmol/L			
ddCTP		15μmol/L		
ddGTP			15μmol/L	
ddTTP				15μmol/L
MgCl <sub>2</sub>	10mmol/L	10mmol/L	10mmol/L	10mmol/L
Tris-HCL pH 7.5	40mmol/L	40mmol/L	40mmol/L	40mmol/L
NaCl	50mmol/L	50mmol/L	50mmol/L	50mmol/L

For each sample, four (1.5ml) microcentrifuge tubes were labelled (A, C, G or T) and 3.5 μl of each respective termination mix were pipetted into the tube. These tubes were then placed in a heat block at 37°C and allowed to equilibrate for several minutes and 4.5μl of isotopically labelled sequencing mix was added. The reactions were allowed to undergo annealing and extension for five minutes at 37°C. The reaction was then stopped by the addition of 5μl of 95% (vol/vol) formamide/20 mM EDTA buffer. When all reactions had been stopped, tubes were stored at 4°C prior to gel electrophoresis (or at -20°C if electrophoresis did not occur within 24 hours). Reactions were run on 6% denaturing (6M urea, 1 x TBE) polyacrylamide gels within 48 hours of

sequencing. Prior to electrophoresis, the sequencing reactions were heat denatured at 95°C for five minutes and immediately placed on ice before being loaded on to the gel.

#### III.4.vii.(b). Polyacrylamide Gel Electrophoresis for Manual Sequencing

Electrophoresis was performed using a vertical gel apparatus system (Hoefer™ Poker Face II). Two clean glass plates, a 37x43cm back plate, and a Sigmacote™ treated (chlorinated organopolysiloxane in heptane, Sigma) 33x41cm top plate were assembled with a bottom spacer and side-spacers (0.4mm) and clamped together to prevent gel leakage prior to polymerisation.

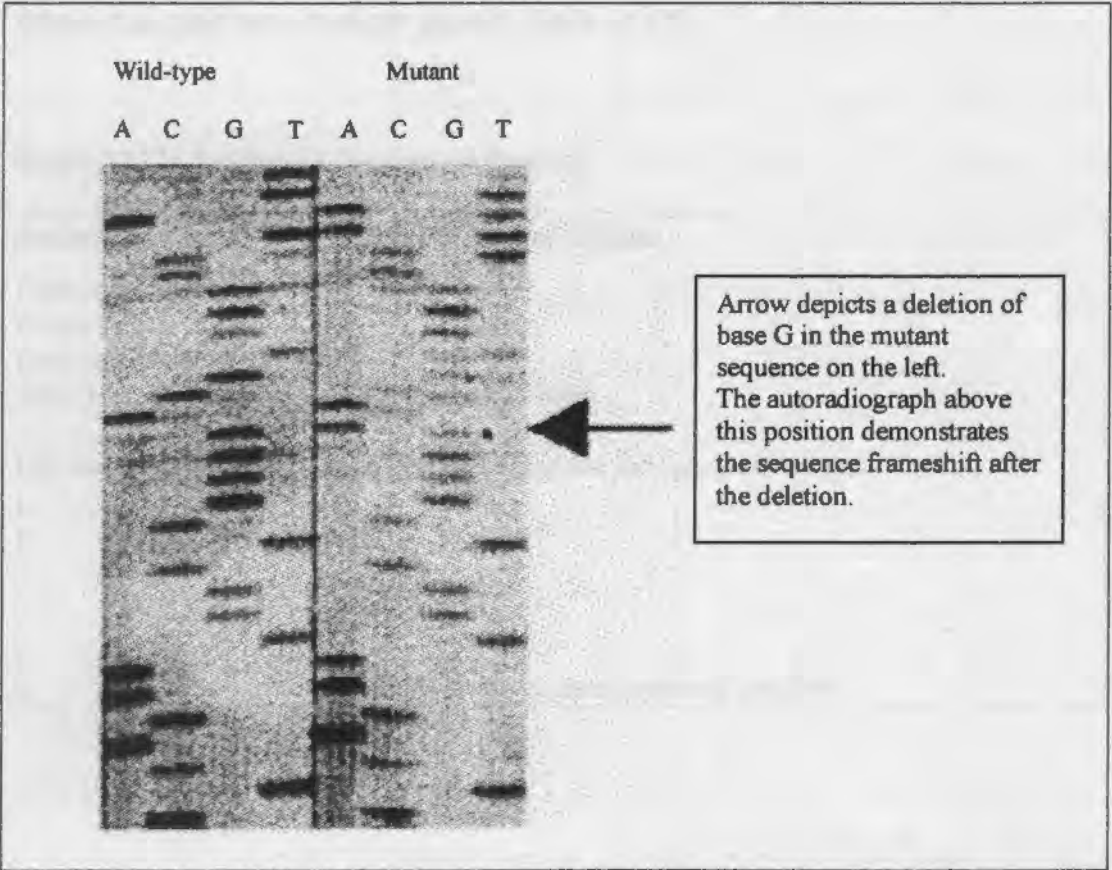
A denaturing 6M urea 6% polyacrylamide (19:1 acrylamide:bis-acrylamide) gel solution (80ml) was prepared with 80 µl of TEMED (Sigma) and 800µl of 10% APS (ammonium persulphate, Sigma). The gel mix was poured between the plates, an inverted sharks-tooth comb was inserted and secured with clamps. The gel was allowed to polymerise for between forty minutes to two hours at room temperature. The clamps, bottom spacer and comb were then removed and the excess gel solution was wiped from the glass. The plates were assembled in the electrophoresis apparatus (Hoefer) with 1XTBE buffer and the gel was pre-warmed at 68W constant power for 30 minutes. Prior to gel loading, excess urea was displaced from the well with buffer and the sharks-tooth comb (64-wells) was inserted to form the loading slots.

Denatured sequencing reactions (2.5µl) were fractionated on 6% denaturing gels at 68W constant power (BIO-RAD PowerPac 3000) for 2-3 hours at 45°C. Bands were

detected by autoradiography after the gel had been fixed and dried on Whatmann paper (3MM).

III.4.vii.(c). Autoradiography of Sequencing Gels.

Dry gels were placed in autoradiographic cassettes and a sheet of X-OMAT autoradiographic paper (Kodak) was layed over the gel. The gels were exposed overnight at room temperature to X-OMAT films (or longer if the isotopic signal was weak) and developed through an automatic X-ray developing machine (Kodak). When developed, the films were examined over a standard white light box and a comparison was made between the resultant and the published sequence (see Fig III.1).



**Fig III.1.** Autoradiograph showing wild-type and mutant sequences

### III.4.viii. Direct Automated DNA Sequencing.

To evaluate the relative efficiency of SSCP, the amplified genomic DNA for *PKD1* exons 44 and 45 from approximately 50 Australian ADPKD patients was screened by direct automated sequencing. Fifteen of these patients had been screened previously by SSCP and the sequence and SSCP results were directly compared. The same previously listed oligonucleotide primers and conditions (Table III.5) were used to prepare templates for sequencing and SSCP analysis. These primers were then used in forward and reverse dideoxy sequencing reactions. Sequencing was performed using the Big-Dye ready reaction terminator cycle sequencing kit (Applied Biosystems Inc. Prism) for electrophoresis on the 377 Sequencer (Applied Biosystems Inc.). The recommended protocol was modified to a total volume of 10 $\mu$ l. Approximately 20ng of clean double-stranded template were used per reaction (Table III.12).

**Table III.12.** Automated Sequencing Protocol.

#### Sequencing Protocol Using Big-Dye Kit (ABI Prism)

Terminator mix	4.5 $\mu$ l
Primer (20ng/ $\mu$ l)	1.0 $\mu$ l
Template (dsDNA)	20ng
ddH <sub>2</sub> O	to a final volume of 10 $\mu$ l

The reaction was then cycle-sequenced using the following conditions:

1.	95°C	5 sec	X 1
2.	95°C	10 sec }	
	50°C	5 sec }	X 25
	60°C	4 min }	
3.	4°C	Hold at this temperature until purified.	

The sequencing reactions were ethanol-salt precipitated as recommended by manufacturer to remove fluorescently labelled dNTPs and dried to zero volume (see Appendix A for details).

Sequencing reactions were mixed with 4 µl of formamide EDTA loading buffer, denatured and electrophoresed in a denaturing (6M urea) 4% polyacrylamide gel (19:1 FMC) on an ABI 377 DNA Sequencer (as per manufacturers recommendation). The sequencing results were analysed using Sequence-Navigator™, a software package by ABI (Perkin-Elmer), or Sequencher 1995 (Gene Codes Inc).

#### III.4.ix. Characterisation of Sequence Variants.

It was necessary to establish whether detected variants were pathogenic or neutral. When a sequence change was detected its genomic position was noted, and if exonic or in the splice-site region, the effects on the mRNA and amino acid sequences were determined. RNA was not available in this study and therefore it was not possible to perform RT-PCR, which means that splice-site mutations could not be fully characterised. Some intronic polymorphisms located in exon recognition sequences can lead to exon skipping, and this type of mutation would remain undetected without RT-PCR.

All sequence variations were analysed using DNAsis (Hitachi Software) or WEBCUTTER (<http://bio.lundberg.gu.se/cutter2/>). These programs also indicate restriction enzyme sites within a given sequence of DNA. When the variant sequence is compared to the published “wild-type” sequence, differences in restriction sites may be

found. The results of these comparisons were used to locate suitable restriction enzymes for family segregation analyses. Restriction digest analyses were also performed on individuals for some common polymorphic changes indicated by SSCP. The use of such restriction digests not only confirms sequence variation results but also negates the need for extra sequencing. The specific restriction enzymes suitable for particular sequence variations are listed in Chapters V & VI. All restriction digests were performed as recommended by the manufacturer.

Digested DNA fragments were run on 4% agarose gels or 10% non-denaturing polyacrylamide gels (19:1) using the Multiphor II Electrophoresis unit (Pharmacia Biotech), which allows rapid separation of fragments. The gels were electrophoresed at a constant temperature of 10°C.

When polymerised, the gels were placed cellophane-side down on the ceramic back-plate of the Multiphor apparatus. Agarose gel buffer strips were prepared (3% agarose pre-soaked in 1xTBE with a few drops of bromophenol blue added), and placed at the upper and lower edges of the gel. The acrylamide gel was then loaded with 2-3 µl of the restriction digest mix. The gels were run at a constant power output of 3 W for between 40 to 60 minutes at 10°C, and then silver stained as previously described (Section III.4.iv). The gels were dried under vacuum on Whatmann paper (3 MM).

If no change in the restriction site was detected, SSCP analysis or sequencing was used for family studies. In addition to family segregation analysis, all missense changes were compared to the wild-type sequence using the secondary structure protein



prediction algorithms PHDsec (Rost *et al.*, 1994) and nnPredict (<http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html>). The wild type and variant sequences were submitted separately and the predicted protein structures were obtained for comparison.

#### III.4.x. Characterisation of Insertions or Deletions.

When sequence analysis indicated an insertion or deletion, a second PCR was carried out on the sample under investigation and the product was cloned using the TOPO TA cloning kit (Invitrogen). The vector with insert was then transformed into TOP10 one shot competent cells. After heat shock at 42°C, the cells were cultured in SOC medium at 37°C, spread on to 50µg/ml ampicillin plates with X-gal and IPTG and incubated overnight at 37°C. On the next day, 8 to 10 white colonies were picked from each plate and grown as separate cultures (4ml of LB medium, 50µg/ml ampicillin) at 37°C overnight. Plasmid DNA was isolated using the QIAGEN Plasmid Mini Kit and then sequenced with M13 Forward (-20) and M13 Reverse primers using the Big-Dye sequencing kit (Applied Biosystems) and a 377 DNA sequencer/genetic analyser.

### III.5. Mutation Detection in the 5' Duplicated Region.

#### III.5.i. Amplification of *PKD1*-specific template in the 5' Duplicated Region

The difficulty associated with amplification of a *PKD1*-unique template in the duplicated region of this gene poses a substantial problem to any mutation screening study. The various methods that were attempted to achieve this specificity are reviewed here. They focus on the long-range amplification of six *PKD1*-specific templates (Table III.15) that in turn are used as template in nested amplification reactions (Tables III.16-21).

A reverse primer in exon 34 similar to those previously described (Watnick *et al.*, 1997; Peral *et al.*, 1997) was designed to act as a *PKD1*-specific anchor for the amplification of exons 23 to 33. Known *PKD1*-specific primer sites in exon 15 (Watnick *et al.*, 1999) and IVS 1 (Thomas *et al.*, 1999) were used to amplify *PKD1* template for exons 15 to 21 and exons 2 to 5. Primer design was based on sequence variations between *PKD1* and a BAC clone (AC002039) containing HG1 and HG2 (Loftus *et al.*, 1999). A comparison between AC002039 and *PKD1* was performed to generate PKD-1 specific primer sites around exon 1. Exon 22 of *PKD1* was not screened. It lies between two polypyrimidine tracts in introns 21 and 22, and as previously reported (Watnick *et al.*, 1997), it is exceedingly difficult to amplify across these tracts in either direction.

BAC and PAC libraries were screened and two clones found to contain HG inserts were partially sequenced. The resulting sequence variations were used to generate specific

oligonucleotide primer sites in IVS11, and XL-template was amplified from that *PKD1*-specific primer site in both forward and reverse directions to include exons 6 to 14.

### III.5.ii. Screening of BAC and PAC DNA Pools for *PKD1* and the HGs

A PCR screen of two human DNA libraries was undertaken (Human BAC DNA pools, Release II, Research Genetics and Human PAC DNA pools, Genome Systems). PAC libraries are similar to BAC libraries and were developed from P1-derived artificial chromosomes which have an average insert size of 130-150 kb (Ioannou *et al.*, 1994). PCR amplifications were performed with primers in both the duplicated and unique regions of the *PKD1* gene, as listed in Table III.13. This experiment aimed to identify clones containing one or more of the homologous genes or the *PKD1* gene. When such clones were detected, the duplicated region was sequenced and analysed to locate differences that could be utilised for *PKD1* unique primer design.

#### III.5.ii(a). Screening BAC and PAC Pools

Both libraries were supplied as “super-pool” and “pool” plates (Research Genetics and Genome Systems) in a 96-well microtitre format. The covering foil or plastic on each plate was removed when frozen (and wiped with 100% ethanol) to prevent cross-contamination between wells, and the microtitre plate contents thawed at room temperature. The PCR screen was first performed on the superpool plate and the addresses of any clones that were successfully PCR amplified using primers for *PKD1* were recorded. The corresponding pool plate and its row and column plates were then amplified with the same primers to identify which individual clones in the superpool

well were positive for the *PKDI* primer sequence. Amplification reactions of 25µl total volume used 1µl of superpool plate DNA as template and 2µl for all other plates.

Initial superpool screening was performed with primers for exons 2 to 3, exon 35 and exon 45. A more detailed screen of the clones was then undertaken using primers for exons 5, 13, 15 and 31-33. The primers, their genomic location, amplification conditions and product size are listed in Table III.13. Cycling conditions include the number of cycles and annealing temperature (the times listed indicate the duration of each repeated step, *eg.* 1'/2'/2' represents 1 minute denaturation, 2 minutes annealing and 2 minutes extension of primers).

**Table III.13.** Primers for DNA Pool Screening

Exons	Primer Sequence	Cycling Conditions	Genomic Position	Product size (bp)
2-3	1810 5' ggaaggcctctggcctacta 3' 1811 3' ggagggcagaagggatattg 3'	55°C (X30) (1'/2'/2')	19848-864 20215-234	387
5	5AF 5' tggagccaggaggagcagaac 3' 5BR 5' cagggcggcaggtgccgctt 3'	65°C (X25) (0.5'/0.5'/1.5')	20771-791 21347-366	595
13	13F 5' tcacctgccacctgggctcac 3' 13R 3' gacaagagcctggtgccacc 3'	65°C (X25) (0.5'/0.5'/1.5')	26279-299 26519-539	261
15	3134 5' gtgctggtcaccagcatcaa 3' 3135 5' cagccagcaggatctgaaaat 5'	55°C (X30) (1'/2'/2')	28507-526 29826-847	1341
31-33	3853 5' ggactcgtccgtgtggac 3' 2930 3' ggatttggcaggcgagtaggggct 5'	68°C (X35) 2-step (0.5'/2')	43672-690 44239-262	591
35	3651 5' ctgcaactgcctctggagg 3' 3652 5' ggatttggcaggcgagtaggggct 3'	68°C (X35) 2-step (0.5'/2')	47330-349 47530-549	220
45	3686 5' cgtcttagctcagctcagct 3' 3687 5' gtgtccctctccccccact 3'	68°C (X35) 2-step (0.5'/2')	50946-965 51322-341	396

### III.5.ii.(b). Culture and Isolation of BAC and PAC DNA

When a BAC or PAC product was identified as positive for *PKDI*, or its HGs, a glycerol stock solution for the specific clone was ordered from Research Genetics or Genome Systems respectively. Stocks were streaked on LB (Luria Bertani) Agar-plates (containing chloramphenicol 12.5µg/ml) using aseptic technique and incubated at 37°C overnight. To obtain sufficient working volumes of the clonal DNA for PCR optimisation and sequencing, a single colony from these plates was picked to inoculate a bacterial culture. Purification of the culture was performed using a modification of the QIAGEN Plasmid Midi Kit protocol (personal communication Dr. Arseni Markoff, Münster) as detailed in Appendix A.

### III.5.ii.(c). Amplification and Sequencing of BAC and PAC DNA

The *PKDI* and HG isolated DNA was then amplified with the primers listed in Table III.14.

**Table III.14.** Primers for BAC and PAC DNA Amplification.

Exons	Primer Sequence	Cycling Conditions	Genomic Position	Product size (bp)
2-5	1810 5' ggaaggcctctggcctacta 3'	68°C (X 5)	19848-864	999
	1807 5' tcagggaggcaggcgacata 3'	2-step (0.5'/2')	20837-846	
7-9	4102 5' cagctacgtctcgcgagctgc 3'	68°C (X 5)	22420-439	791
	4103 5' ggtgagggaaggcttcacggc 5'	2-step (0.5'/2')	23191-210	
9-10	4104 5' gcaggtgtaccggctcctca 3'	68°C (X 5)	23261-270	454
	4105 5' cagctgggtcctgtgtgccg 3'	2-step (0.5'/2')	23695-714	
11-12	4106 5' gaagaccatgtccgagccgg 3'	68°C (X 5)	25086-105	945
	4107 5' gaagaccatgtccgagccgg 3'	2-step (0.5'/2')	26011-030	
12-14	4108 5' cgcggcgggtcttcaagctctg 3'	68°C (X 0)	26099-119	945
	4140 5' gagggtgttgggaggaagg 3'	2-step (0.5'/2')	27023-043	

After the PCR amplification of each clonal DNA with these primers, the products were electrophoresed in a 1% agarose gel (a total PCR volume of 25 $\mu$ l was loaded into wells formed with a wide tooth comb). Bands were excised from the gel and purified using the QIAquick Gel Extraction Kit as recommended by the manufacturer (QIAGEN). The concentration of DNA after cleaning was estimated against DNA of known concentration on an agarose gel. Approximately 30ng of DNA were then sequenced in both the forward and reverse directions for each amplicon using the individual primers listed in Table III.14.

Sequencing was performed using the ready reaction terminator cycle sequencing kit, as recommended by the manufacturer (Applied Biosystems Inc. Prism). Clean pellets were air-dried and then sent to the Institute of Physiology, WWU, Münster, Germany where samples were electrophoresed on an ABI 373 sequencer by the designated operator (Appendix A).

#### III.5.ii.(d). *PKD1*-Specific Sequence in the Duplicated Region of the Gene.

Comparisons between the resultant BAC and PAC sequences and the published sequence of *PKD1* (GenBank Acc. L39891) were made by Dr. Arseni Markoff using the GeneStream global alignment tool (<http://www2.igh.cnrs.fr/home.eng.html>). *PKD1* introns 3, 4, 8, 9, 12, 13 and 14 were assessed for areas specific to *PKD1*. On the basis of these results (Chapter IV), *PKD1* specific primers were designed in IVS11 to amplify *PKD1* exons 6 to 11 and 12 to 14. These are listed in Table III.15 with the other *PKD1* primers used for XL-PCR amplification.

**Table III.15. Primers for XL-Template Amplification.**

Exons	Primer Sequence	Cycling Conditions	Genomic Position	Product size (bp)
Ex 2-5	IVS1 F 5' cccagcgtctcatctgtctgg 3' IVS5 R 5' ggtgggaacgagggtgtcaac 3'	68°C (X40) 2-step (0.5'/5')	17049-69 21579-90	4,542
Ex 6-11	2927 5' gccgcctacagcatcgtggccc 3' 11R 5' cagtgaggggaggcacctacact 3'	68°C (X40) 2-step (0.5'/4')	21447-470 25622-644	4,198
Ex 12-14	12F 5' cagtgtaggtgcctcccctcactg 3' 3130 5' ccagcacggtcaggaggtact 3'	68°C (X40) 2-step (0.5'/2')	25621-644 27410-430	1810
Ex 15	IVS13F 5' ctgtcccggttcactcactgc 3' Ex15XLR 5' caacgtgggcctccaagtagttg 3'	68°C (X40) 2-step (0.5'/4')	26769-89 30607-29	3861
Ex 15-21	Ex15XLF 5' gcaactacttgaggcccacg 3' IVS21R 5' gcaggggtgagcaggtggggccatc 3'	68°C (X40) 2-step (0.5'/4')	30606-26 33957-80	3375
23-34	TWF1* 5' ctgcactgacctcacgcatgt 3' Ex34R 5' atgtgggtgtcttgggtagg 3'	68°C (X40) 2-step (0.5'/6')	37674-694 44403-422	6,749

The amplification reactions were performed using long-range amplification enzyme with a hot-start (XL-rTth Applied Biosystems).

To ensure that diluted long-range PCR products were *PKD1*-specific, a series of dilutions ( $1:10^{-2}$ ,  $1:10^{-3}$  and  $1:10^{-5}$ ) were made prior to re-amplification and used to amplify exon 8 of *PKD1*. A dilution factor of  $\geq 1:10^{-4}$  of PCR product had previously been shown to produce *PKD1*-specific amplification when used as template (Watnick *et al.*, 1999). Diluted PCR products were then purified from an excised agarose gel slice using the QIAquick PCR Purification Kit as per the manufacturer's recommendations (QIAGEN). The purified products were sequenced using the SequiTherm EXCEL II DNA Cycle Sequencing Kit, as described by the manufacturer (Epicentre Technologies), and products were isotopically labelled with [ $\alpha$ - $^{33}\text{P}$ ]-dATP (Hartmann Analytics). The detailed protocol is listed in Appendix A.

Prior to electrophoresis, the products of the sequencing reactions were denatured at 95°C for 3 minutes and immediately placed on ice. Denaturing PAGE electrophoresis and autoradiography was performed as previously described (Section III.5.vii). The sequencing results (Chapter IV) indicate that a dilution of  $1:10^5$  produced specific amplification of the *PKD1* gene, whereas dilutions of  $1:10^2$  and  $1:10^3$  gave rise to the co-amplification of homologous genes.

### III.6. Mutation Detection Screen of *PKD1* Exons 2 to 33

#### III.6.i. PCR Amplification of individual *PKD1* Exons

Genomic DNA was amplified with XL primers (Table III.15), and the products were diluted  $1:10^5$ . The diluted products (1µl in total reaction volume of 25µl) acted as template for the nested PCR amplification of individual exons contained within the respective XL-PCR products (III.6.i.a-g). Each PCR reaction included 200µM of each dNTP, 1.5mM MgCl<sub>2</sub>, 30ng of forward and reverse primer, 1 X Buffer, 1U of *Taq* polymerase and 5µl of Q-solution (QIAGEN).

Despite repeated attempts at the optimisation of reaction conditions we were unable to produce a PCR product of suitable quality for mutational analysis of the *PKD1* gene surrounding exon 1. The high GC content of the region surrounding exon 1 prevented the specific amplification for screening purposes. Other groups have reported the difficulty of trying to screen for mutations in exon 1 of the *PKD1* gene (Rossetti *et al.*, 2001; Phakdeekitcharoen *et al.*, 2001). Both groups were only able to screen the region after performing long-range amplifications, followed by nested amplifications using



high annealing temperatures and PCR enhancing additives (DMSO buffer with Taq extender and Advantage-GC genomic polymerase and GC melt respectively).

III.6.i.(a) Amplification of exons 2 to 5

**Table III.16.** Primers for the PCR Amplification of *PKD1* Exons 2-5.

Exons	Primer Sequence	Cycling Conditions	Genomic Position	Product size (bp)
Ex 2	Ex2F 5' ggaaggcctctggcctacta 3' Ex2R 5' ggggattcggcaaagctgat 3'	64°C (X35) (0.5'/0.5'/1')	19845-64 20064-74	230
Ex 3	Ex3F 5' ctacgtgtgggggattcca 3' Ex3R 5' gg cagaagggatattggggg 3'	58°C (X35) (0.5'/1'/1')	20045-64 20211-30	186
Ex 4	Ex4F 5' ggctggcatagacccttccc 3' Ex4R 5' cctggctgggaaggacagag 3'	64°C (X35) (0.5'/0.5'/1')	20378-97 20630-50	273
Ex 5(a)	Ex5aF 5' gtggagccaggaggagca 3' Ex5aR 5' gggaagacgtgctggagg 3'	58°C (X35) (0.5'/1'/1')	20770-87 21081-98	329
Ex 5(b)	Ex5bF 5' cctccagcacgtcttccc 3' Ex5bR 5' ccgttccacctgcacgt 3'	64°C (X35) (0.5'/0.5'/1')	21081-98 21335-52	272
Ex 5(c)	Ex5cF 5' ctacgcctgctggggacaga 3' IVS5 R 5' ggtgggaacgagggtgtcaac 3'	64°C (X35) (0.5'/0.5'/1')	21315-35 21579-89	275

### III.6.i.(b) Amplification of exons 6 to 11.

**Table III.17.** Primers for the PCR Amplification of *PKD1* Exons 6-11.

Exons	Primer Sequence	Cycling Conditions	Genomic Position	Product size (bp)
6	4349 5' gtgagtgtctgtgccca 3' 4350 5' cctcctcctcctgagact 3'	58°C (X30) (1'/1'/1')	21491-508 21868-886	396
7	4351 5' ggctctgagcctcagttt 3' 4352 5' taaccacagccagcgtct 3'	58°C (X30) (1'/1'/1')	22104-121 22476-493	391
8	4353 5' gtctgttcgtcctggtgt 3' 4354 5' ccatttcactgggcaca 3'	58°C (X30) (1'/1'/1')	22575-592 22863-880	306
9	4345 5' gttcgggtaggggagttct 3' 4346 5' gtgaaagctcagagaggcca 3'	60°C (X 35) (0.8'/0.8'/0.8')	23090-108 23341-360	271
10	4347 5' ggtggcctgtgggcaaatca 3' 4348 5' gcctgaggagatgcaggga 3'	61°C (X30) (1'/1'/1')	23608-627 23929-947	340
11-11a	11AF 5' ggggtccacgggccatgaccgt 3' 11AR 5' cgcagtcacagggttgggcctc 3'	66°C (X25) (0.5'/1'/1')	24300-320 24610-630	331
11-11b	11BF 5' acggaacagctcaccgtgtct 3' 11BR 5' aagggtggccaccagggcaggg 3'	66°C (X25) (0.5'/1'/1')	24581-600 24880-900	320
11-11c	11CF 5' gcagtgctcagcggcctttg 3' 11CR 5' agcacctgtctgcaggcacc 3'	66°C (X25) (0.5'/1'/1')	24849-869 25150-170	321

Exon 11 was amplified as three fragments (a, b & c) due to its size (756bp).

### III.6.i.(c) Amplification of exons 12 to 14.

**Table III.18.** Primers for the PCR Amplification of *PKD1* Exons 12-14.

Exons	Primer Sequence	Cycling Conditions	Genomic Position	Product size (bp)
12	4355 5' ccaggaggcgacaggcta 3' 5135 5' cctggccctgattggcgtc 3'	61°C (X30) (0.8'/0.8'/0.8')	25923-940 26256-74	352
13	5228 5' tcacctgccacctgggtcac 3' 5136 5' ccgaggctcagaaagca 3'	68°C (X35) 2-step (0.5'/2')	26279-299 26581-598	320
14	5002 5' ctgtcccgggttcactcactgc 3' 4140 5' gagggctgttggggaggaagg 3'	68°C (X35) 2-step (0.5'/2')	26769-789 27023-043	275

### III.6.i.(d) Amplification of exon 15

**Table III.19.** Primers for the PCR Amplification of *PKD1* Exon 15

Exon 15 Frag.	Primer Sequence	Cycling Conditions	Genomic Position	Product size (bp)
16	16F 5' cgggtggggagcaggtgg 3' 16R 5' gctctgggtcaggacagggga 3'	68°C (X35) 2-step (0.5'/2')	27351-67 27610-30	280
15	15F 5' cgcttgggggtttcttt 3' 15R 5' acgtgatgttgcgcccg 3'	53°C (X30) (0.5'/0.5'/1')	27569-86 27821-38	270
14	14F 5' gccccctgggtggtcagc 3' 14R 5' caggctgcgtgggatgc 3'	68°C (X35) 2-step (0.5'/2')	27790-807 28022-39	250
13	13F 5' ctggagggtcgtcgcgtt 3' 13R 5' ctggctccacgcagatgc 3'	67°C (X30) (0.5'/0.5'/1')	27991-008 28229-46	259
12	12F 5' cgtagaacaggcgcatca 3' 12R 5' gcagcagagatgttgtggac 3'	58°C (X30) (0.5'/0.5'/1')	28203-20 28452-72	270
11	11F 5' ccaggtcctatcttgtgaca 3' 11R 5' tgaagtccactgtgctgtgt 3'	62°C (X30) (0.5'/0.5'/1')	28433-53 28661-81	249
10	10F 5' ctacctgtgggatctgggg 3' 10R 5' tgctgaagctcacgctcc 3'	65°C (X30) (0.5'/0.5'/1')	28602-20 28801-18	217
9	9F 5' gggctcgtcgtcaatgaag 3' 9R 5' caccacctgcagcccctcta 5'	66°C (X30) (0.5'/0.5'/1')	28758-77 29005-34	277
8	8F 5' ccgcccaggacagcatcttc 3' 8R 5' cgctcccagcatgttg 3'	66°C (X30) (0.5'/0.5'/1')	28966-85 29208-25	260
7	7F 5' cggcaaaggcttctcgtc 3' 7R 5' ccgggtgtggggaagctatg 3'	63°C (X30) (0.5'/0.5'/1')	29147-65 29414-33	287
6	6F 5' cgagccatttaccacccatag 3' 6R 5' gccagcaccagctcacat 3'	58°C (X30) (0.5'/0.5'/1')	29398-418 29610-28	231
5	5F 5' ccacgggcaccaatgtgag 3' 5R 5' ggcagccagcaggatctgaa 3'	66°C (X30) (0.5'/0.5'/1')	29598-616 29829-48	251
4	4F 5' cagcagcaaggttggtggc 3' 4R 5' gcgtaggcgacccgagag 3'	63°C (X30) (0.5'/0.5'/1')	29791-808 30106-23	333
3	3F 5' acgggcactgagaggaacttc 3' 3R 5' accagcgtgctgttctact 3'	66°C (X30) (0.5'/0.5'/1')	30065-85 30251-70	206
2	2F 5' gccgcagctcacctagac 3' 2R 5' tcggccctgggctcatct 3'	66°C (X30) (0.5'/0.5'/1')	30177-95 30424-41	265
1	1F 5' ctgggactttggggatgggt 3' XL15R 5' caacgtggcctccaagtagttg 3'	63°C (X30) (0.5'/0.5'/1')	30388-407 30607-29	242

Primers first described by Watnick *et al.*, 1999.

### III.6.i(e) Amplification of exons 15 to 21.

**Table III.20.** Primers for the PCR Amplification of *PKD1* Exons 15-21

Exons	Primer Sequence	Cycling Conditions	Genomic Position	Product size (bp)
15(a)	Ex15XLF 5' gcaactacttggaggcccacg 3' Ex15aR 5' ctcaatgatgggcaccaggcg 3'	66°C (X35) (0.5'/0.5'/1')	30606-26 30872-92	287
15(b)	Ex15bF 5' catccaggccaatgtgacggt 3' Ex15bR 5' cctggtggcaagctgggtgtt 3'	64°C (X35) (0.5'/0.5'/1')	30841-61 31086-106	266
16	Ex16F 5' ctggatgggctctcagg 3' Ex16R 5' cctggtggcaagctgggtgtt 3'	56°C (X30) (0.5'/0.5'/1')	31197-214 31468-85	289
17	Ex17F 5' gggccccccagtccttcag 3' Ex17R 5' ccatccccagccgcccaca 3'	66°C (X35) (0.5'/0.5'/1')	32262-81 32489-508	247
18	Ex18F 5' gcccctcaccaccccttct 3' Ex18R 5' gatcccgctgcctccccca 3'	66°C (X35) (0.5'/0.5'/1')	32571-90 32895-914	344
19	Ex19F 5' gtgatgccgtgggaccgtc 3' Ex19R 5' gtgagcaggtggcagtctcg 3'	66°C (X35) (0.5'/0.5'/1')	32946-65 33213-32	287
20	Ex20F 5' ctgctcaccacccctctg 3' Ex20R 5' ggtcccaagcacgcatgca 3'	64°C (X35) (0.5'/0.5'/1')	33225-43 33445-63	239
21	Ex21F 5' cgctgctgacagcttgctgtgcc 3' IVS21R5' gcagggtgagcaggtggggccatc 3'	66°C (X30) (0.5'/0.5'/1')	33763-86 33957-80	218

### III.6.i.(f) Amplification of exons 23 to 33

**Table III.21.** Primers for the PCR Amplification of *PKD1* Exons 23-34.

Exons	Primer Sequence	Cycling Conditions	Genomic Position	Product size (bp)
23(a)	TWF1* 5' ctgcactgacctcagcatgt 3' 23a-R 5' caccagaaagatgggctgcac 3'	58°C (X30) (0.5'/0.5'/1')	37674-694 38006-026	353
23(b)	23bF 5' ggggccctggccaacctcag 3' 23bR 5' cgtgtgccccaccgctgca 3'	68°C (X35) 2-step (0.5'/2')	37979-998 38336-355	377
24	24F 5' tgtgacctgcgcttctg 3' 24R 5' ccaggctggcccgagag 3'	68°C (X35) 2-step (0.5'/2')	38600-617 38786-803	204
25	25F 5' ctgggctcagctccgctacc 3' 25R 5' cctcgactctgcagaggctc 3'	68°C (X35) 2-step (0.5'/2')	38927-946 39236-255	329
26	26F 5' cggctctatcctgagaaggc 3' 26R 5' cagcacagccagtgagagca 3'	68°C (X35) 2-step (0.5'/2')	39272-291 39567-586	315
27	27F 5' cctcccacccctccctcttg 3' 27R 5' cagagcttggcagggtccgc 3'	68°C (X35) 2-step (0.5'/2')	41000-019 41205-224	225
28	28F 5' cgagcctgacctccctcctg 3' 28R 5' cggagtgggaccatggaacg 3'	58°C (X30) (0.5'/0.5'/1')	41254-273 41438-457	204
29	29F 5' cgtggcctcctgcagtgcgg 3' 29R 5' ggagggaagaggctgccccg 3'	68°C (X35) 2-step (0.5'/2')	41486-505 41751-770	285
30	30F 5' cacctgtgtggcctcctctc 3' 30R 5' ggctccattcccagttactcc 3'	68°C (X35) 2-step (0.5'/2')	41794-813 41973-992	199
31	31F 5' gctgaccactgcctcgtc 3' 31R 5' agtccaagctgcgccaagg 3'	68°C (X35) 2-step (0.5'/2')	43585-603 43773-791	207
32	32F 5' ttggcgcagcttgactc 3' 32R 5' cagggtcgcaggtttctc 3'	58°C (X30) (0.5'/0.5'/1')	43775-792 43879-896	123
33	33F 5' taccttgcctgacccgcgc 3' 33R 5' ggtgagcttcagagcccc 3'	68°C (X35) 2-step (0.5'/2')	44064-081 44296-313	250

Exon 23 was amplified as two fragments (a & b) due to its size (630bp).

\* Previous reference Watnick *et al.*, 1997.

### III.6.ii. Purification and Sequencing of Nested PCR Products

5µl aliquots of the 25µl nested PCRs were electrophoresed on 1% agarose gels in 1X TBE at 100 V for 40-60 minutes, stained with ethidium bromide and photographed under UV-transillumination as previously described (Section III.4.ii). When amplicons resulted in a clearly detectable PCR product, the remaining 20µl were purified through QIAquick PCR spin columns (QIAGEN), and eluted in 20µl of ddH<sub>2</sub>O. This is a lower elution volume than recommended by the manufacturer, however, sufficient template was recovered to proceed with direct automated sequencing reactions.

### III.7. Investigation of *PKD1* Homologous Gene Sequences

In addition to specific mutation detection of *PKD1* in the duplicated region sequence, information on the BAC and PAC clones was also used to investigate the possible functions of the homologous genes. This work was undertaken as a direct result of the present study by Dr. Nadja Bogdanova and Dr. Arseni Markoff (Institute for Human Genetics, Münster, Germany) using the following methodology.

#### III.7.i. Reverse Transcription and PCR (RT-PCR)

Total RNA was isolated from confluent astrocytoma T98G human cell cultures using the Trizol™ reagent (Gibco BRL) according to the manufacturer's protocol. This reagent is based on the single-step phenol/guanidine isothiocyanate isolation protocol originally described by Chomczynski and Sacchi (1987). The RNA was subjected to reverse transcription, primed with random hexamers and either *PKD1*- or *PKD1* homologue- specific primer as detailed below.

### Total RNA Isolation Protocol

1. 1ml of astrocytoma cell suspension was divided between two 1.5ml microcentrifuge tubes and centrifuged at 6,000 rpm for 5 minutes at 4°C.
2. The supernatant was discarded and the pellet was resuspended in 750µl of Trizol reagent (2ml of reagent per  $1 \times 10^6$  cells). The tubes were shaken to homogenise cells.
3. RNA was isolated in an aqueous phase from DNA and organic compounds by the addition of 0.2 volumes of chloroform to the homogenate. The tubes were shaken for 15 seconds and incubated at room temperature for 2-3 minutes.
4. The homogenate was then centrifuged at 12,000g for 15 minutes at 4°C.
5. The RNA-containing upper-phase was removed into clean tubes and equal volumes of isopropanol used to precipitate the RNA.
6. The tubes were then incubated at -20°C for 10 minutes and centrifuged for a further 10 minutes at 4°C to pellet the RNA.
7. The supernatant was removed, the RNA pellets were washed with 500µl of 75% ice-cold ethanol and the tubes centrifuged at 7,500 rpm for 5 minutes at 4°C.
8. The ethanol was removed and the pellets air-dried for 15-20 minutes. A paper tissue was placed over tubes to prevent air-borne contamination from entering the tubes.
9. The pellets were resuspended in 30µl of RNase-free ddH<sub>2</sub>O at 55°C for 10 minutes.
10. The quality of RNA and the possibility of DNA or RNase contamination was assessed by the electrophoresis of a 5µl aliquot sample on a 2% Agarose gel stained, with ethidium bromide.

Translationally active RNA was prepared from the polysomal fraction (membrane-bound polysomes) as previously described (Mechler, 1987). The microsomal fraction was obtained through sucrose gradient centrifugation of a T98G cytoplasmic extract and mRNA was separated and purified from the bound polysomes. When the RNA was considered to be of sufficient quality, cDNA was produced by reverse transcription using Superscript II reverse transcriptase (Gibco BRL Life Technologies) as detailed below.

### Reverse Transcription

Total RNA and mRNA (0.5 - 3µg) were primed with random hexamer (100pM) and 3' oligonucleotide primer (10pM) (5' GCCATCCCCGAAGGTCCAGTCGAA 3'), which was common to *PKD1* and the homologous genes. RNase-free water was then added to a total volume of 12.5µl. The mix was incubated at 65°C for 10 minutes and immediately placed on ice. An enzyme mix was prepared from 1µl of Superscript II reverse transcriptase (Gibco BRL Life Technologies), 5µl of 5 x RT 1<sup>st</sup> strand buffer, 2.5µl of DTT (0.1M), 3µl of 10mM dNTPs (Gibco BRL Life Technologies), and 1µl of RNase inhibitor (Gibco BRL Life Technologies), in a final volume of 12.5µl. The RNA and enzyme preparations were mixed and incubated at the following temperatures, 25°C for 10 minutes, 42°C for 60 minutes, 50°C for 30 minutes and 70°C for 10 minutes.

The first round of PCR amplification was performed on 1µl aliquots of the reverse transcribed products with primers listed in Table III.22. The PCR product of this first reaction (1µl) was then used as template in a second amplification using *PKD1* and HG specific primers (see Table III.22). Agarose gel electrophoresis (Section III.4.ii) of



cDNA was then performed to assess if the homologous genes produced translationally active mRNA after specific amplification (see Chapter IV).

**Table III.22.** Primers for Amplification of *PKD1* and Homologous Genes

Exon	Primer Sequence	Cycling Conditions	Genomic Position	cDNA product size
5	Forward outer primer 5' gtgtcccgctcctcgggtgcaga 3'	40 cycles of 94°C 45sec 55°C 1 min 72°C 4 mins	21373-94	2,387 bp
15	Reverse outer primer 5' ccgaagtcacacgtgtaaagaa 3'		27581-602	
11	Forward <i>PKD1</i> nested primer 5' cttgaggcccaaccctggact 3'	40 cycles of 94°C 45sec 68°C 1 min 72°C 1 min	24607-27	715bp
13	Reverse <i>PKD1</i> nested primer 5' ccggcactgtggagacctgcaga 3'		26395-417	
11	Forward HG nested primer 5' cgtgaggcc-aaccctgggct 3'	40 cycles of 94°C 45sec 68°C 1 min 72°C 1 min	24607-27	(approx) 715bp
13	Reverse HG nested primer 5' ctggcactgtagagaccggcagg 3'		26395-417	

#### **IV. Results:**

##### **Specific Amplification of the Duplicated Region of *PKD1***

## IV. Results: Specific Amplification of the Duplicated Region of *PKD1*

### IV.1. Introduction

The existence of genes highly homologous to *PKD1* raises interesting biological questions related to their physiological role, and at the same time presents a major obstacle to the specific amplification of the *PKD1* gene and hence to the study of the molecular defects leading to ADPKD. During the course of this study, the genomic sequence of two of the homologues, HG1 and HG2, contained in a single BAC clone (Accession # AC002039) became publicly available (Loftus *et al.*, 1999). However, earlier studies aimed at the cloning of *PKD1* (The European PKD Consortium, 1994; The International PKD Consortium, 1995; Hughes *et al.*, 1995), had demonstrated the existence of three homologous transcripts, suggesting that at least one additional homologous gene remained to be characterised. To address the questions of the sequence characteristics of the homologues, their role, and the selection of *PKD1*-specific amplification primers, a collaboration as indicated below was conducted. The genomic sequences of human genomic clones containing *PKD1* homologues, or the *PKD1* gene itself, were compared and their expression was examined.

The work on the characterisation of the *PKD1* homologous genes was undertaken in collaboration with Dr. Nadja Bogdanova and Dr. Arseni Markoff at the Institute for Human Genetics, Westfälische-Wilhelm's University, Münster, Germany. The candidate acknowledges their contribution and presents the joint results with their full knowledge and approval.

#### IV.2. Identification of human genomic clones containing *PKDI* and its homologous genes

As described in Chapter III (Section III.5.ii), two human genomic libraries, Human BAC DNA Pools (Release II, Research Genetics Inc.) and Human PAC DNA Pools (Genome Systems), were screened by PCR analysis for clones containing inserts of the duplicated and unique regions of *PKDI*. Two initial PCR amplifications were performed with primers for *PKDI*. The first experiment amplified exons 2 and 3 in the duplicated region of the gene. This amplification aimed to detect clones containing *PKDI*, and clones containing *PKDI* homologous genes. The second amplification was performed with primers for exon 45 in the unique region of the *PKDI* gene; this experiment was performed to distinguish between *PKDI*-containing and HG-containing clones.

Screening of the BAC library super-pool identified two wells that contained clones with inserts corresponding to the duplicated region of *PKDI*. Subsequent pool plate screening confirmed the presence of two clones with unique addresses (plate addresses 204-o-15 and 267-d-9), that were PCR-positive for exons 2-3, 5, 13, 15, and 31-33 of *PKDI*. Attempts to amplify the unique region (exons 35 or 45) of *PKDI* in these clones were unsuccessful, indicating that the inserts corresponded to homologous genes, rather than to *PKDI*. Partial sequence analysis of the homologous genes contained in these clones revealed that they were not identical to other identified *PKDI* homologues, HG1 and HG2 (Loftus *et al.*, 1999) see Chapter VII, and they were named HG3 and HG4.

Screening of the PAC library identified two clones (80-I-8 and 136-M-14), which were PCR positive for exons 2 and 3 located in the duplicated region and exon 45 in the unique region of *PKD1*. These results indicated that the two PAC clones contained the *PKD1* gene (see Figure IV.1).

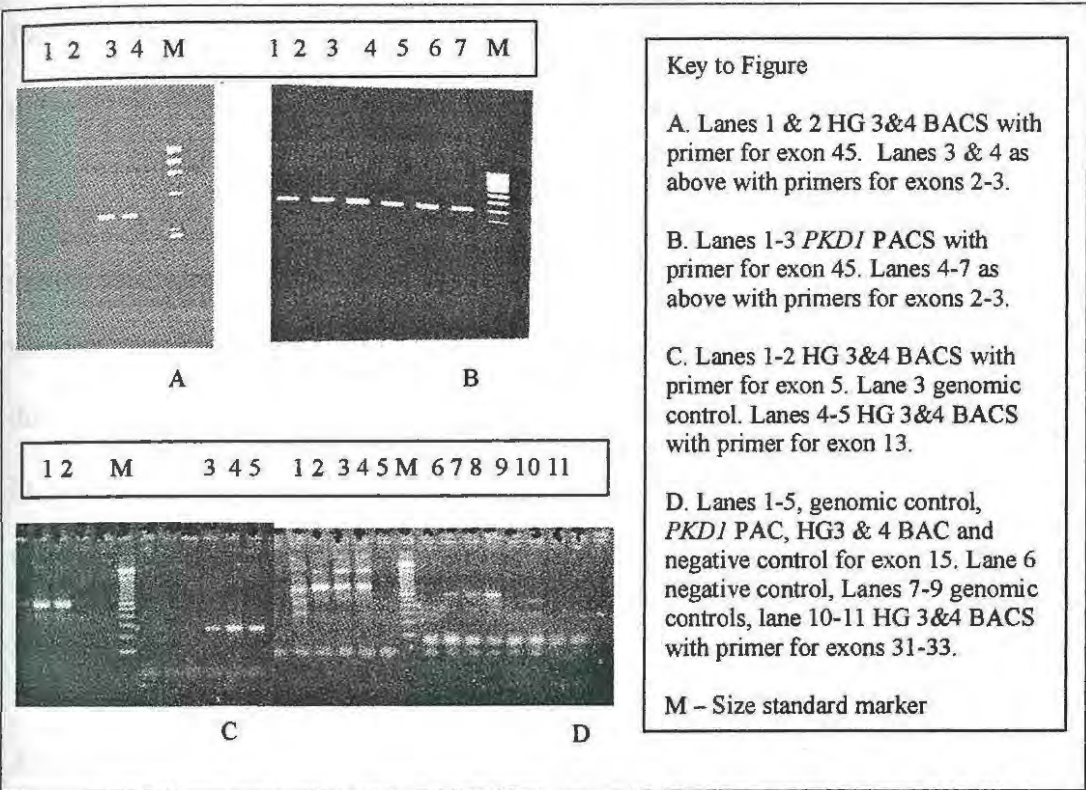


Fig. IV.1. Amplification results of BAC and PAC clones.

The glycerol stocks of the *PKD1*- and HG-containing clones were purchased from the manufacturers, propagated and BAC and PAC DNA were extracted and purified for sequence analysis.

### IV.3. Sequence analysis of *PKD1* and the homologous genes

To identify *PKD1*-specific primers and amplify specific templates for mutation detection analysis, the two clones containing the homologous genes and the *PKD1*-

containing clone were sequenced and compared. This study attempted to characterise a region of *PKD1* that had not been screened previously. Early studies had screened *PKD1*-specific templates in the duplicated region by amplifying from exon 34 in the unique region up to exon 23 (Peral *et al.*, 1997; Watnick *et al.*, 1997). Two subsequent studies reported screening for mutations in exons 15 to 21 (Watnick *et al.*, 1999b; Thomas *et al.*, 1999). For that reason, the present study focused on the sequence 5' of exon 15 of *PKD1*.

The region from exon 2 to IVS14 (nucleotide positions 19,866 to 27,595 of the *PKD1* genomic sequence Accession # L39891) was amplified, using as templates DNA from the two BAC clones (204-o-15 and 267-d-9) that contain the HGs, and from one of the PAC clones (136-M-14) representing *PKD1*. The region was covered by five PCR fragments. These PCR products were analysed by automated direct sequencing as described in chapter III.

The resulting sequences have been deposited in GenBank under Accession numbers AF320593 and AF320594 (Bogdanova *et al.*, 2001).

Sequence alignment, performed by Dr. Markoff with the GENESTREAM alignment tool (<http://www.2.igh.cnrs.fr/home.eng.html>), showed that the two sequences derived from BAC clones 204-o-15 and 267-d-9 differed from each other, as well as from the previously published HG1 and HG2 (Loftus *et al.* 1999). Compared to *PKD1*, the sequence of the newly identified homologous genes, HG3 and HG4, contained small deletions or insertions (5 to 21 base pairs), and nucleotide substitutions. The

sequence identity for this 7.7kb region was found to be 96.4% between *PKD1* and HG3 and 96.5% between *PKD1* and HG4.

Further analysis of GenBank, using the BLAST homology search tool at NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>), revealed two additional recent entries (Accession numbers AC010488 and AC040158) which are highly homologous to *PKD1* and to its homologous genes. These findings point to the existence of a greater number of homologous genes, at least six, than suggested in early publications (The European PKD Consortium, 1994; The International PKD Consortium, 1995; Hughes *et al.*, 1995).

#### IV.4. Functional characteristics of the homologous genes

The functional importance of the *PKD1* homologous genes was investigated in the glioblastoma cell line T98G by RT-PCR. The presence of *PKD1*, HG3, and HG4 transcripts was examined in total T98G RNA, and in translationally active RNA prepared from a polysomal cell fraction. RT-PCR for exon 12, located in the duplicated region and thus present in all three genes, was performed with primers specific to *PKD1* or to HG3 and HG4. Total RNA yielded a product for exon 12 with both *PKD1* and the homologous gene primers. By contrast, in the translationally active RNA, only a *PKD1*-specific product was detected (see Appendix, Bogdanova *et al.*, 2001). This experiment indicates that whilst the HGs are transcribed, they are not found in the translationally active polysomal fraction and therefore are unlikely to be translated. These findings are supported by the sequence analysis of the coding regions of HG3 and HG4. The sequence was found to contain a number of premature

stop codons, suggesting that even if the HGs were translated they would generate short, and most probably non-functional, peptides (Bogdanova *et al.*, 2001).

On the basis of these results, it was concluded that the *PKD1* homologous genes HG3 and HG4 are highly likely to be pseudogenes. We have proposed to refer to these genes as *PKD1* pseudogenes 3 and 4, *PKDIP3* and *PKDIP4* (Bogdanova *et al.* 2001).

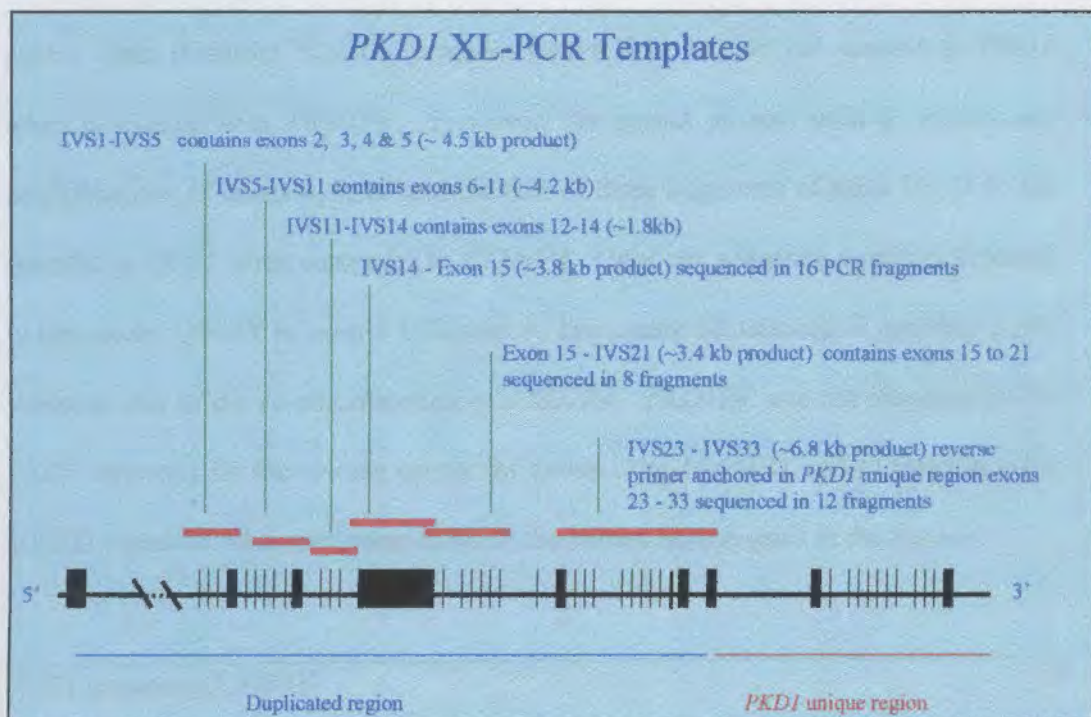
#### IV.5. Design of *PKD1*-Specific Oligonucleotides

We have attempted to improve the specificity of PCR amplification and diagnostic sequence analysis of *PKD1* by selecting primers for long-range amplification (XL-PCR) that are specific to *PKD1* in the duplicated region of the gene (exons 2 to 14), when compared to all its known homologous genes.

All XL-PCR primers used (Fig IV.2), are specific to *PKD1* relative to the *PKDIP1* (HG1) and *PKDIP2* (HG2) sequences published in a previous study (Loftus *et al.*, 1999). We have also compared the *PKD1* sequence between exons 2 to 14 to that of *PKDIP3*, *PKDIP4* and *PKDIP5* (Bogdanova *et al.*, 2001). *PKDIP6* was not included in this analysis, since its sequence homology to *PKD1* in this region extends over only 2530 bp of the *PKD1* gene, corresponding to *PKD1* exon 2 to exon 7.



## *PKD1* XL-PCR Templates



**Fig IV.2.** XL-PCR primers used for the amplification of *PKD1* (Exons 2-33)

To amplify a *PKD1*-specific product we chose a novel primer site in IVS 11 that was specific to *PKD1* when compared to the sequence of the homologous gene. The site was specific to *PKD1* compared with *PKDIP1*, *PKDIP2*, *PKDIP3* and *PKDIP5*. The fourth homologue, however, was found to be identical to the *PKD1* gene in this primer site region. The primer site in intron 11 was utilised in the design of forward (12F) and reverse (11R) primers as shown in Figure IV.3. Primer 6F in IVS 5 was used in an XL-PCR amplification reaction with primer 11R to amplify a region of 4.2 kb, containing exons 6 to 11 (Fig IV.2), using a long-range polymerase enzyme (XL-rTth enzyme, Applied Biosystems). Forward primer 12F with a primer in IVS 14 (14R) (Fig IV.2) was used to produce a 1.8 kb *PKD1*-specific template, including exons 12 to 14.

The fourth homologue was found to be identical to the *PKD1* gene in the 6F and 11R primer sites, therefore XL-PCR products (exons 6 to 11) are not specific to *PKD1* when compared with *PKDIP4*. However, the nested primers used in subsequent amplifications of exons 6, 9, 10 and the first of three fragments of exon 11 (11A) are specific to *PKD1* when compared to *PKDIP4*. Only one sequence variation detected in this study, Q563X in exon 8 (Chapter V, Truncating Mutations), is possibly a HG variation due to the co-amplification of *PKDIP4*. *PKDIP4* was not identical to the *PKD1* sequence for the reverse primer for exons 12 to 14 (14R), and so this fragment is *PKD1*-specific when compared to all of the known homologues in the region.

*PKD1* sequence (L39891).

PKD1 Primer 12F	5'-CAGTGTAGGTGCCTCCCCTCACTG 3' (g.25621-44)
AC002039 PKD1P1	5' CAGTGTAGGTGCTGTCCCTGTGAG 3' (g.27954-78)
AC002039 PKD1P2	5' CAGTGTAGGTGCTGTCCCTGTGAG 3' (g.67648-71)
AF320593 PKD1P3	5' CAGTGTAGGTGCTGTCCCTGTGAG 3' (g.5762-5787)
AF320594 PKD1P4	5' CAGTGTAGGTGCCTCCCCTCACTG 3' (g.5786-5809)
AC010488 PKD1P5	5' <u>TGAGGCCCG</u> TGCCTCCCCTCACTG 3' (g.59954-78)
AC040158 PKD1P6	5' Not present in this region

PKD1 Primer 11R    5'CAGTGAGGGGAGGCACCTACACT 3' (g.25622-44)

Reverse Primer 11R is positioned at the same primer site as 12F, so it shares the same homology as depicted above.

**Fig IV.3.** Alignment of *PKD1* and HG Sequences in the IVS 11 Primer Region

Exons 2 to 5 were amplified using a previously published *PKD1*-specific forward primer site in IVS1 (Thomas *et al.*, 1999), and a reverse primer in exon 6.

IV.6. Specificity of the PCR amplification

To confirm the specificity of the nested PCR technique with our newly designed primers, the nested PCR products amplified from the XL-PCR template were sequenced. The initial XL-PCR amplification of exons 6 to 11 was chosen as the test fragment, this was followed by the nested amplification of exon 8, where the template was used at various dilutions ( $1:10^{-4}$ ,  $1:10^{-3}$  and  $1:10^{-2}$ ). The resulting sequences were compared for co-amplification of *PKD1* and its homologous genes (Figure IV.5). The experiments showed that a dilution factor greater than  $1:10^{-4}$  was necessary to produce a *PKD1*-specific template from long-range PCR amplified product.

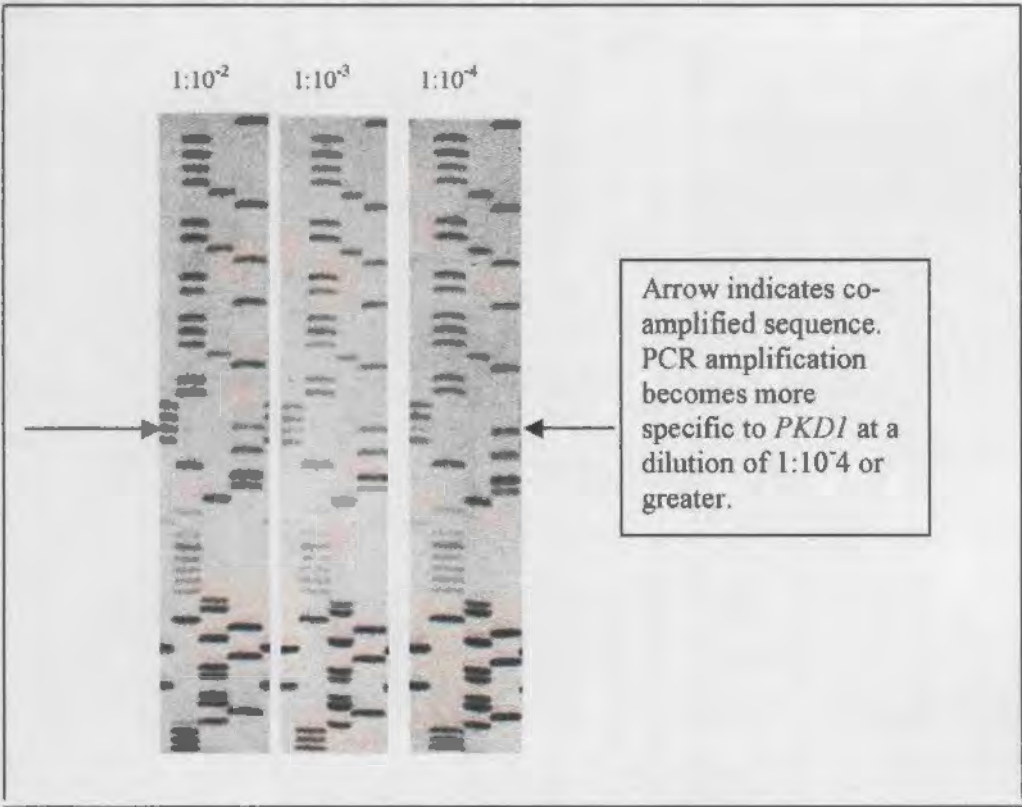


Fig IV.4. Co-amplification of *PKD1* & HGs in nested PCR.

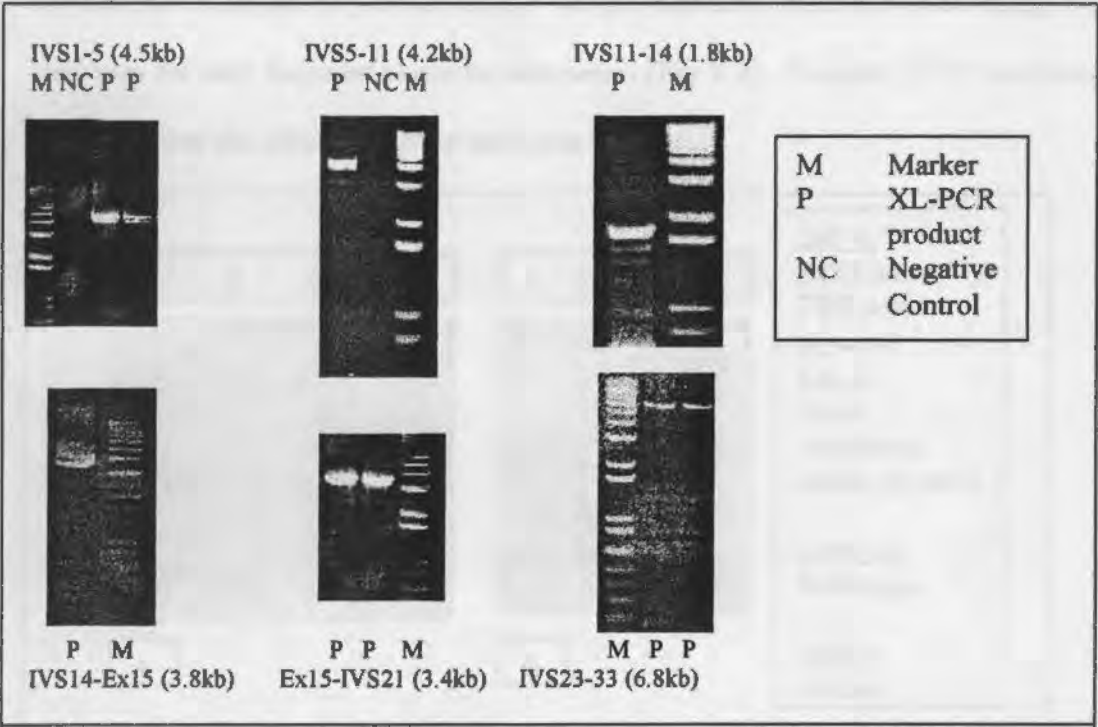
As indicated in Chapter III (Section III.6), we used a previously reported primer site in exon 15 (Watnick *et al.*, 1999), specific to *PKD1* when compared to HG1 and HG2

(Loftus *et al.*, 1999), to amplify two long-range fragments. The first fragment included most of exon 15 (IVS14F & 15R), and the second amplified the rest of exon 15 to exon 21 (15F & IVS21R). The sequence of *PKDP3*, *PKDP4* and *PKDP5* has yet to be determined for this region, and therefore a comparison between these homologues and the *PKDI* sequence was not possible. An anchored primer in the unique region of *PKDI* with a forward primer in the duplicated region was used to amplify exons 23 to 33, as described previously (Watnick *et al.*, 1997).

## **V. Results – Mutations**

**Chapter V. Introduction**

The mutation analysis of *PKD1* spanned 94% of the coding sequence, namely exons 2-21 and 23-33 in the duplicated region, and exons 34-41 and 43-46 in the unique part of the gene. The duplicated region was investigated using *PKD1*-specific primers, previously published or designed in this study as a result of the findings described in Chapter IV. These primers were used for the long-range PCR amplification of *PKD1* fragments ranging in size from 1,810 bp to 6,749 bp, followed by nested PCR amplification and direct sequencing of the resulting products. This analysis was performed on 17 unrelated patients from Australian ADPKD families showing “positive” linkage to the *PKD1* region on chromosome 16 (see p.93).

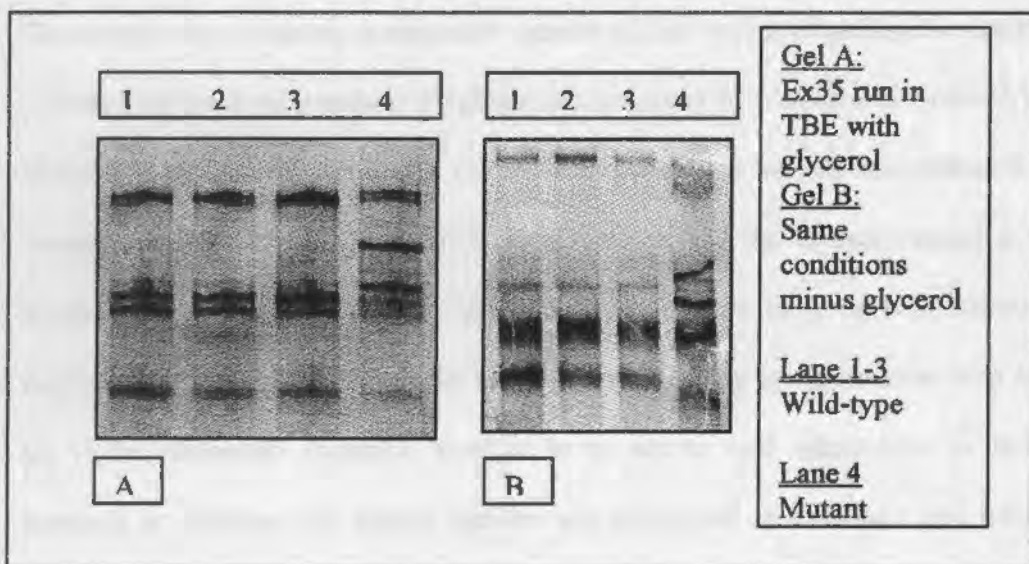


**Fig V.1** Results of XL-template amplification in the duplicated region of *PKD1*



The study of the unique 3' region of *PKDI* was performed in two stages: a) screening by SSCP analysis, which detects conformational changes and electrophoretic mobility shifts resulting from sequence variation; b) specifying the nature of the sequence variants by direct sequencing analysis. The mutation analysis of the unique *PKDI* region was performed in 48 unrelated individuals from ADPKD families: 15 Australian, 26 Bulgarian and 7 Polish all showing "positive" linkage to the *PKDI* region on chromosome 16 (as described p.93).

Initial PCR amplification and subsequent SSCP analysis were optimised for each of the 13 individual fragments used to screen the unique region of the gene. The most appropriate PCR amplification, SSCP gel composition, and electrophoretic conditions are detailed in Chapter III (Section III.4). The gel conditions were optimised using PCR templates for each fragment as can be seen below (Fig V.2). Variable SSCP conditions markedly alter the efficiency of the technique.



**Fig V.2.** SSCP optimisation for mutation screening

The efficiency of the SSCP screening was assessed in the 17 Australian ADPKD patients by re-analysing part of the unique region, namely exons 44 and 45, and fragment A of exon 46. These exons were selected on the basis of the different size of the PCR fragments, which is a factor known to affect SSCP sensitivity (Orita, 1989), as well as on the basis of the large number of disease-causing mutations detected by other groups in this region of *PKD1* (Turco *et al.*, 1995 & 1997; Neophytou *et al.*, 1996; Rossetti *et al.*, 1996 & 2001; Roelfsema *et al.*, 1997; Daniells *et al.*, 1998a & b; Aizal *et al.*, 1999; Perrichot *et al.*, 1999).

The direct sequencing approach was also applied to 4 additional Australian patients from *PKD1*-linked families that had joined the study at a later stage, as well as to 31 affected individuals where linkage analysis had not been possible due to small family size.

The criteria for assigning a sequence variant to the group of pathogenic mutations followed, as much as possible, the guidelines proposed by Cotton and Scriver (1998). The following criteria were met: (1) the same sequence variant was detected upon repeated analysis starting from PCR amplification; (2) the change caused a direct truncation of the protein either through a nucleotide substitution generating a premature stop codon or through a shift in the reading frame leading to a premature stop codon; (3) if the nucleotide change(s) resulted in an amino acid substitution or in-frame insertion or deletion, the altered residue was conserved in evolution and computer analysis of secondary protein structure suggested conformation changes; (4) a family study (performed whenever possible) indicated that the sequence variant segregated



with the disease phenotype. In some cases, family members had provided fingerprick blood samples collected on to a filter paper card rather than whole blood. Due to the amount of genomic template required for long-range PCR amplification, segregation analysis in those few families could not be performed.

The analysis resulted in the detection of a total of 74 sequence variants, of which 24 were classified as probable disease-causing mutations (Table V.1). Twenty-one of the proposed disease-causing mutations, including all 12 mutations in the duplicated region of the gene, have been discovered as a result of this study.

The mutation analysis included a total of 19 Australian patients with evidence of linkage to chromosome 16. The probable pathogenic mutation was identified in 14 out of 19, giving a 74% overall mutation detection rate for this group.

Mutation detection in the *PKDI* gene is technically and financially demanding and is therefore most efficient when preceded by linkage analysis. This pre-condition was met for the core group of 19 Australian patients where *PKDI* was systematically screened. Nonetheless, a less demanding analysis of part of the unique region of the gene was conducted in 31 affected subjects with unknown linkage status. This part of the study was productive, with three mutations identified in this additional group of patients.

**Table V.1.** Summary of probable disease-causing mutations in the *PKD1* gene.

Exon or IVS	Affected Family ID	Sequence Variation	Predicted effect	Possible Confirmatory Assay
Ex 5	A9	c.1352G>T g.21430G>T	Missense G381C	Sac II abolished
Ex 8	A4	c.1898C>T g.22716C>T	Nonsense Q563X	none
IVS 14	A22	IVS14+1G>C g.26942G>C	Splicing Variant	Alu I created
Ex 15	A52	c.5105-5106del (AT) g.29004delAT	Frame-shift deletion 2bp I1632fsX1656	Dde I created
Ex 15	A36	c.5159delG g.29058delG	Frame-shift deletion 1bp V1650fsX1721	Afl III created
Ex 15	A61	c.6764T>G g.30662T>G	Missense Y2185D	Ava II created
Ex 15	A15	c.7017C>G g.30915C>G	Nonsense S2269X	Bsm AI created
IVS 16	A5	IVS16+1G>T g.31394G>T	Splicing Variant	Mae II created
Ex 18	A46	c.7472-74del (ACC) g.32650-52del	In-frame deletion 2421delT	none
Ex 23	A6	c.8565G>A g.37893G>A	Missense G2785D	none
Ex 25	A42	c.9289-9327del g.39090del39	In-Frame deletion 3027del13 aa	Visible on 3% agarose gel with ethidium bromide
Ex 28	A10	c.9847delC g.41352delC	Frame-shift deletion 1 bp F3213fsX3315	Bsl I created
Ex 34	B22	c.10635del 23 g.44374del 23	Frame-shift P3475fsX3617	Ava II abolished
IVS 37	B57	c.IVS37-10 C>Ag.48491C>A	Splicing mutation fsX3684	Bsr I created

Ex 40	B77	c.11587delG g.49515delG	Frame-shift deletion 1bp R3793fsX3824	None
Ex 40	P101	c.11520C>A g.49448C>A	Nonsense S3770X	Eag I abolished
Ex 41	B25	c.11669-11674del g.49737-49742del	In-frame deletion 3820-3821del QE	Alu I abolished
Ex 44	A18.1 & A41.1	c.12218C>T g.50791C>T	Nonsense Q4003X	Mae I (or Bfa I) created
Ex 44*	A72	c.12239 C>T g.50812C>T	Nonsense Q4010X	Mae I created
Ex 44**	A64.1	c.12269C>T g.50842C>T	Nonsense R4020X	Blp I created
Ex 44***	A3	c.12332 C>T g.50905C>T	Nonsense Q4041X	Mae I created
Ex 45	A54	c.12347C>A g.51003C>A	Missense L4046F	BbvI abolished
Ex 46	A38	c.12891 G>C g.51637G>C	Missense R4227P	None
Ex 46	B41	c.13069ins39 g.51815ins39	In-frame insertion a.a.4286ins13	None

\*Previous reference Daniells *et al.*, 1998b. \*\*Previous reference Rossetti *et al.*, 1996.

\*\*\*Previous reference Turco *et al.*, 1995.

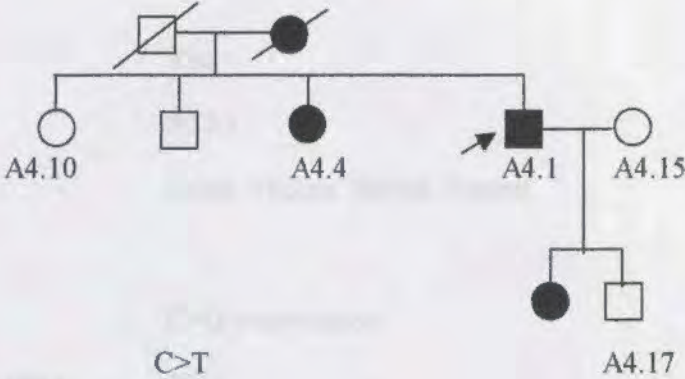
The remainder of this chapter deals with the individual disease-causing mutations in *PKD1*. The results are classified according to the type of molecular change: protein truncating, splice-site variation, missense and non-truncating insertions and deletions. As previously noted, some mutations may not be verified as pathogenic due to lack of RNA for RT-PCR. Mutations are named according to the recommendations of Cotton & Scriver (1998), however we have used traditional numbering for the nucleotide position in the gene (*i.e.* the first encoded nucleotide of the cDNA is at position 212 [L33243] and position 3648 [L39891] of the genomic DNA).

## TRUNCATING MUTATIONS

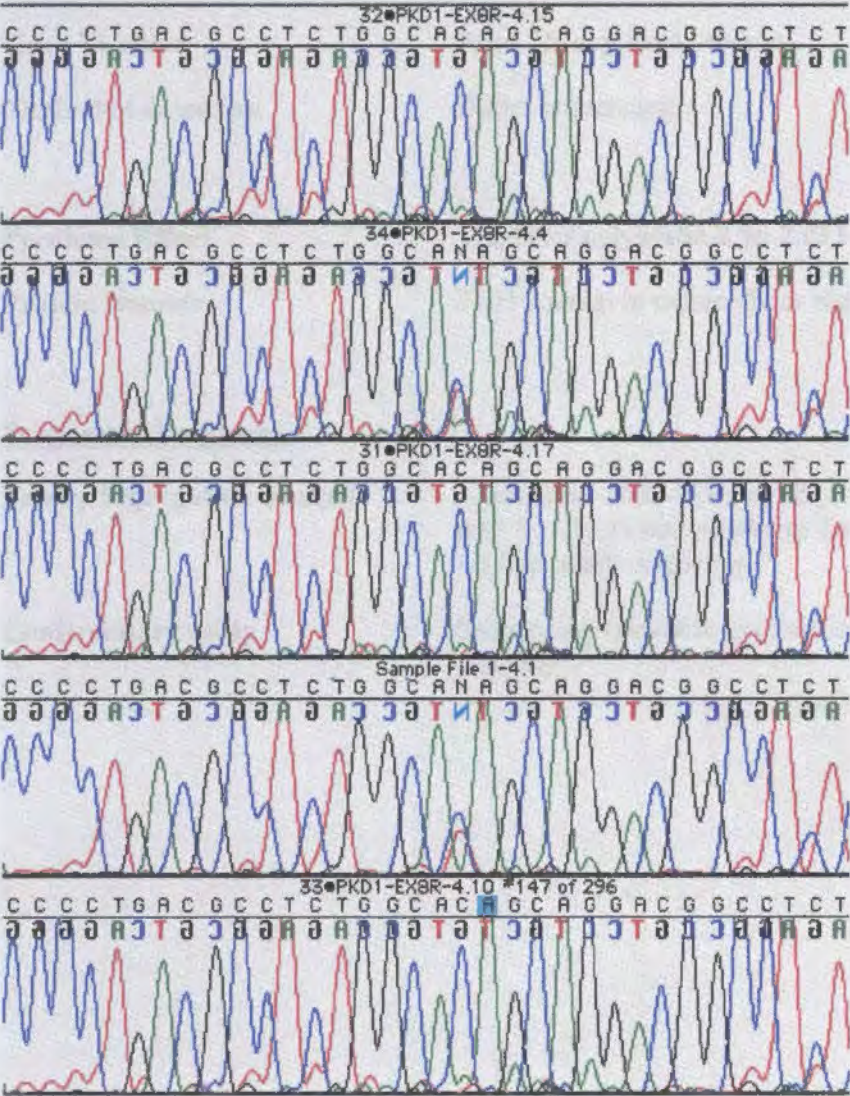
### **Mutation – Q563X**

<b>Novel</b>	Yes	
<b>Patient</b>	A4.1	
<b>Ethnic Origin</b>	Australian, British descent	
<b>Nucleotide Change</b>	C>T transition	
<b>Position</b>	<b>c.DNA</b>	1898
	<b>g.DNA</b>	22716
<b>Gene Location</b>	Exon 8, duplicated region	
<b>Method of detection</b>	Direct sequencing	
	NB. This mutation was detected in a PCR fragment which may be co-amplified with PKD1P4 (see Chapter IV).	
<b>Predicted Effect</b>	Truncates polycystin-1 by 3,739 aa	
<b>Protein Domain</b>	Extracellular region, downstream of C-type lectin domain	
<b>Supporting Evidence:</b>		
<b>Family Segregation Analysis</b>	Segregates with disease in 2 affected (A4.1, A4.4) and absent in 3 unaffected (A4.10, 4.15 and 4.17) family members	
<b>Confirmatory assay</b>	Repeat sequencing (no restriction enzyme assay possible)	

**Pedigree:**



**Electrophoretogram:**



A4.15
A4.4
A4.17
A4.1
A4.10

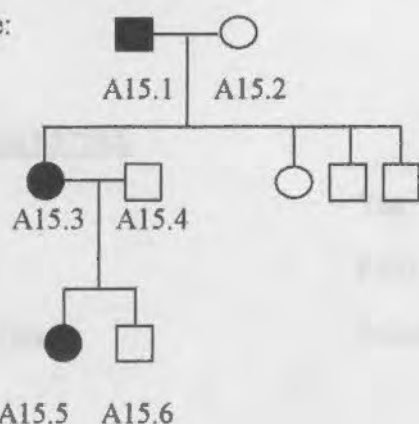
**Mutation – S2269X**

<b>Novel</b>	Yes
<b>Patient</b>	A15.1
<b>Ethnic Origin</b>	South African, British descent
<b>Nucleotide Change</b>	C>G transversion
<b>Position</b>	<b>c.DNA</b> 7017
	<b>g.DNA</b> 30915
<b>Gene Location</b>	Exon 15, duplicated region
<b>Method of detection</b>	Direct sequencing
<b>Predicted Effect</b>	Truncates polycystin-1 by 2,033 aa
<b>Protein Domain</b>	PKD domain in extracellular region

**Supporting Evidence -**

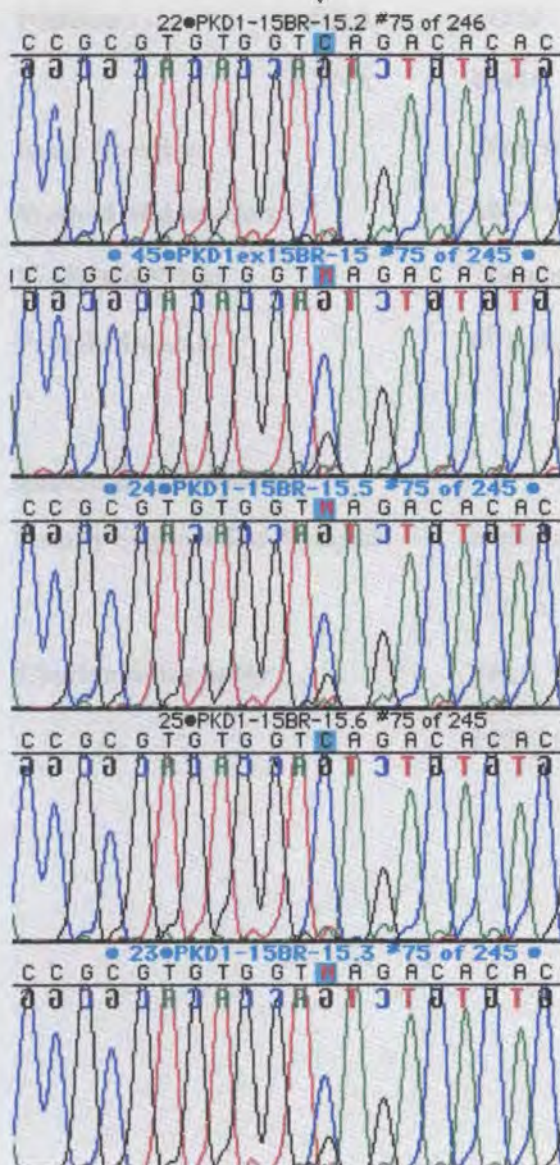
<b>Family Segregation Analysis</b>	Segregates with disease in 3 affected (A15.1, A15.3, A15.5) and absent in 2 unaffected (A15.2, A15.6) family members
<b>Confirmatory assay</b>	Sequencing (possible enzyme assay - Bsm AI site created)

Pedigree:



Electrophoretogram:

↓ C>G



A15.2

A15.1

A15.5

A15.6

A15.3



**Mutation S3770X**

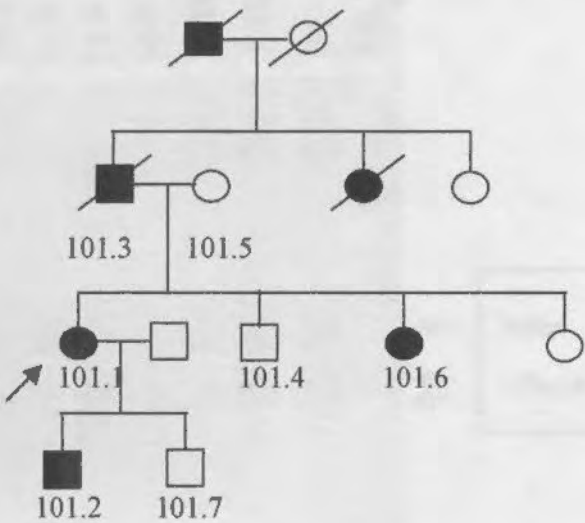
<b>Novel</b>	Yes	
<b>Patient</b>	P101.1	
<b>Ethnic Origin</b>	Polish	
<b>Nucleotide Change</b>	C>A transversion	
<b>Position</b>	<b>c.DNA</b>	11520
	<b>g.DNA</b>	49448
<b>Gene Location</b>	Exon 40, unique region	
<b>Method of detection</b>	SSCP and sequencing	
<b>Predicted Effect</b>	Truncates polycystin-1 by 533 aa	
<b>Protein Domain</b>	6 <sup>th</sup> trans-membrane (TM) domain	

**Supporting Evidence -**

<b>Family Segregation Analysis</b>	Segregates with disease in 4 affected (101.1, 101.2, 101.3 and 101.6) and absent in 3 unaffected (101.5, 101.4 and 101.7) family members
<b>Confirmatory assay</b>	Eag I restriction enzyme assay – restriction site abolished (264 bp in mutant allele and 145bp + 119bp in wild-type)



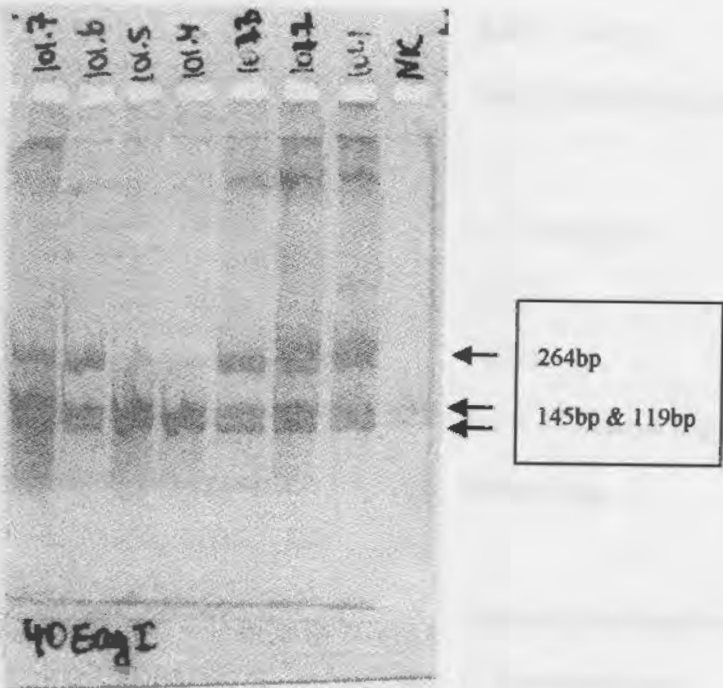
**Pedigree:**



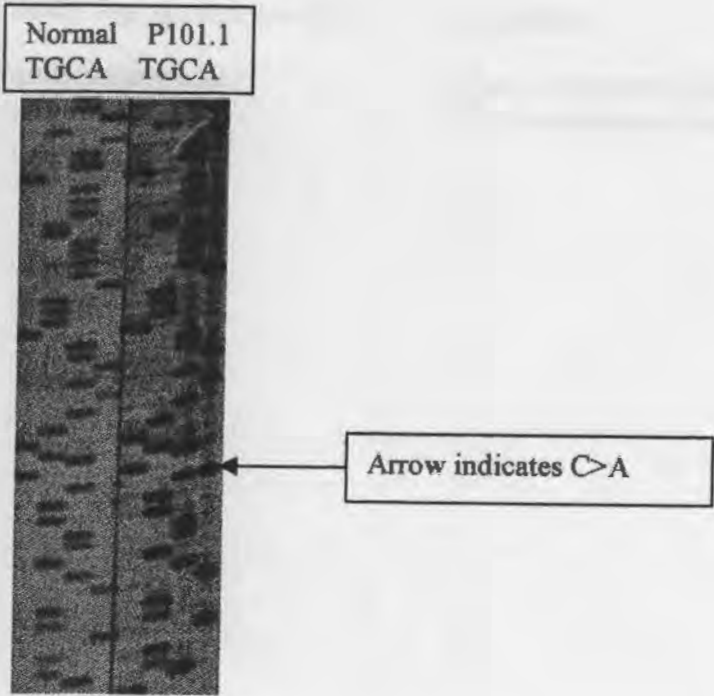
**SSCP gel:** Shows SSCP analysis of DNA from family P101 Exon 40



**Restriction digest of Exon 40 for Family P101 with EagI**



**Sequencing Gel S3770X**



**Mutation Q4003X**

<b>Novel</b>	Yes
<b>Patients</b>	A18.1 & A41.1
<b>Ethnic Origin</b>	Both Australian of Macedonian descent

<b>Nucleotide Change</b>	C>T transition
--------------------------	----------------

<b>Position –</b>	<b>c.DNA</b>	12218
	<b>g.DNA</b>	50791

<b>Gene Location</b>	Exon 44, unique region
----------------------	------------------------

<b>Method of detection</b>	Sequencing
----------------------------	------------

<b>Predicted Effect</b>	Truncates polycystin-1 by 299 aa
-------------------------	----------------------------------

<b>Protein Domain</b>	Extracellular loop
-----------------------	--------------------

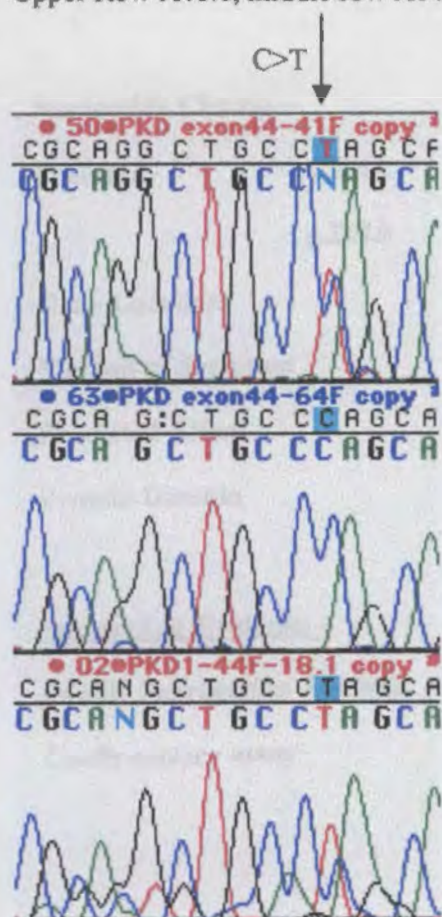
**Supporting Evidence -**

<b>Family Segregation Analysis</b>	Not possible
------------------------------------	--------------

<b>Confirmatory assay</b>	Direct sequencing (Possible confirmatory assay - Mae I restriction site created)
---------------------------	--

### Electrophoretogram:

Upper Row A41.1, middle row A64.1 (wild-type), lower row A18.1



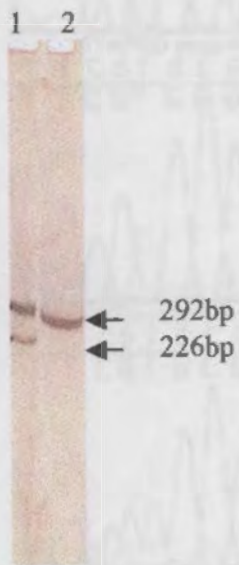
## **Mutation – Q4010X**

<b>Novel</b>	No: Daniells <i>et al.</i> , 1998b & Kim <i>et al.</i> , 2000b.	
<b>Patient</b>	A72.1	
<b>Ethnic Origin</b>	Australian, British descent	
<b>Nucleotide Change -</b>	C>T transition	
<b>Position –</b>	<b>c.DNA</b>	12239
	<b>g.DNA</b>	50812
<b>Gene Location</b>	Exon 44, unique region	
<b>Method of detection</b>	SSCP and direct sequencing	
<b>Predicted Effect</b>	Truncates polycystin-1 by 293 aa	
<b>Protein Domain</b>	Between 8 <sup>th</sup> and 9 <sup>th</sup> TM domains	
<b>Supporting Evidence -</b>		
<b>Family Segregation Analysis</b>	Not possible	
<b>Confirmatory assay</b>	Mae I restriction enzyme assay (restriction site created - 226 bp + 66 bp in mutant allele and 292 bp in wild-type)	

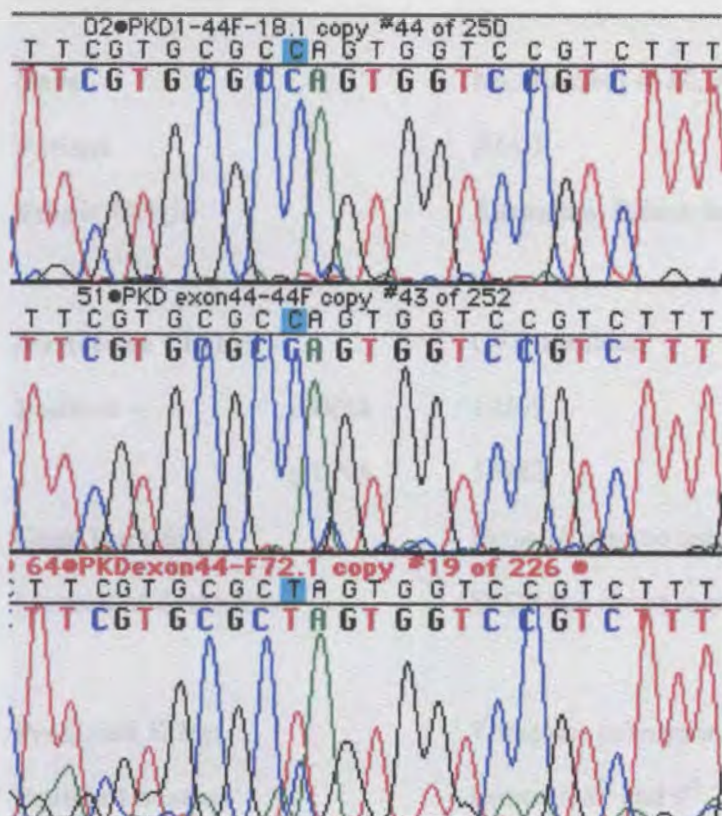
## Restriction Assay Mae I

Lane 1 A72.1

Lane 2 normal control



# Sequencing Gel



A18.1 (wild-type)

A44.1 (wild-type)

A72.1 (C>T Q401X)

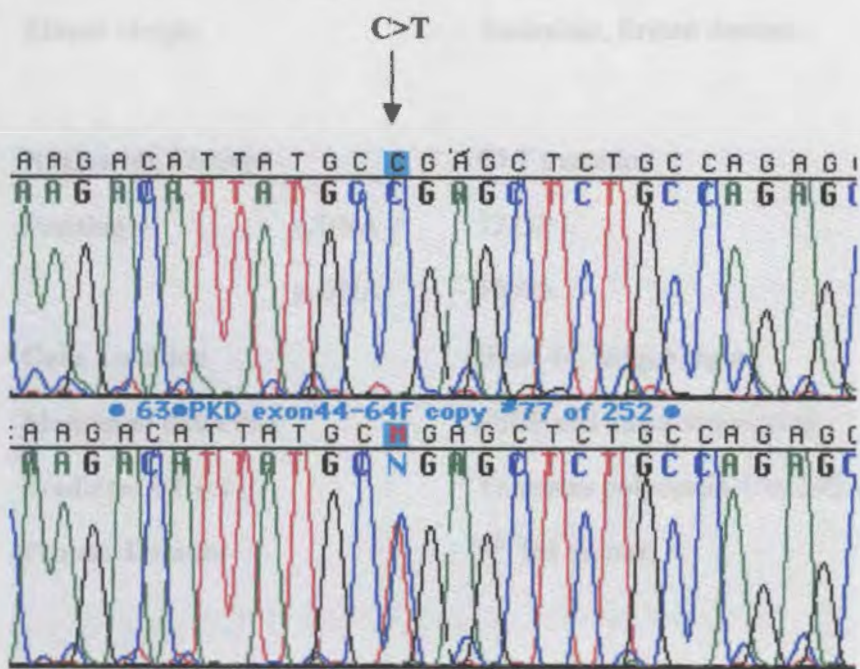
## **Mutation – R4020X**

<b>Novel</b>	No: Rossetti <i>et al.</i> , 1996
<b>Patient</b>	A64.1
<b>Ethnic Origin</b>	Australian, British descent
<b>Nucleotide Change -</b>	C>T transition
<b>Position –</b>	
<b>c.DNA</b>	12269
<b>g.DNA</b>	50842
<b>Gene Location</b>	Exon 44, unique region
<b>Method of detection</b>	SSCP and direct sequencing
<b>Predicted Effect</b>	Truncates polycystin-1 by 282 aa
<b>Protein Location</b>	Between 8 <sup>th</sup> and 9 <sup>th</sup> TM domains
<b>Supporting Evidence -</b>	
<b>Family Segregation Analysis</b>	Not possible
<b>Confirmatory assay</b>	Direct sequencing (possible restriction enzyme assay – Bln I site created)



## Electrophoretogram:

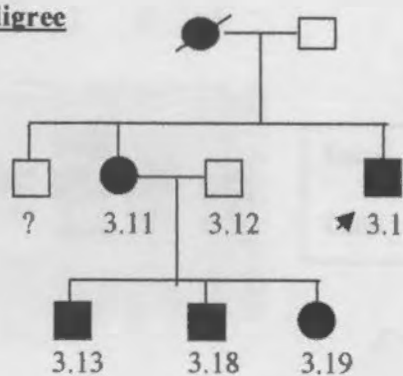
Upper row normal control; lower row A64.1



## **Mutation – Q4041X**

<b>Novel</b>	No: Turco <i>et al.</i> , 1995; Daniells <i>et al.</i> , 1998b Torra <i>et al.</i> , 1998; Rossetti <i>et al.</i> , 2001.	
<b>Patient</b>	A3.1	
<b>Ethnic Origin</b>	Australian, British descent	
<b>Nucleotide Change -</b>	C>T transition	
<b>Position –</b>	<b>c.DNA</b>	12332
	<b>g.DNA</b>	50905
<b>Gene Location</b>	Exon 44, unique region	
<b>Method of detection</b>	SSCP and direct sequencing	
<b>Predicted Effect</b>	Truncates polycystin-1 by 262 aa	
<b>Protein Domain</b>	9 <sup>th</sup> TM domain	
<b>Supporting Evidence -</b>		
<b>Family Segregation Analysis</b>	Segregates in 5 affected (A3.1, A3.11, A3.13, A3.18, A3.19), not found in 1 unaffected (A3.12) family members.	
<b>Confirmatory assay</b>	Mae I restriction enzyme assay (restriction site created - 159 bp + 133 bp in mutant allele and 292 bp in wild-type)	

## Pedigree



## Restriction Assay: Mae I

Lanes :

- 1 A3.1
- 2 A3.11
- 3 A3.12
- 4 A3.13
- 5 A3.18
- 6 A3.19
7. normal control
8. size standard 100bp ladder (Gibco BRL) – no well shown on gel.

1 2 3 4 5 6 7 8

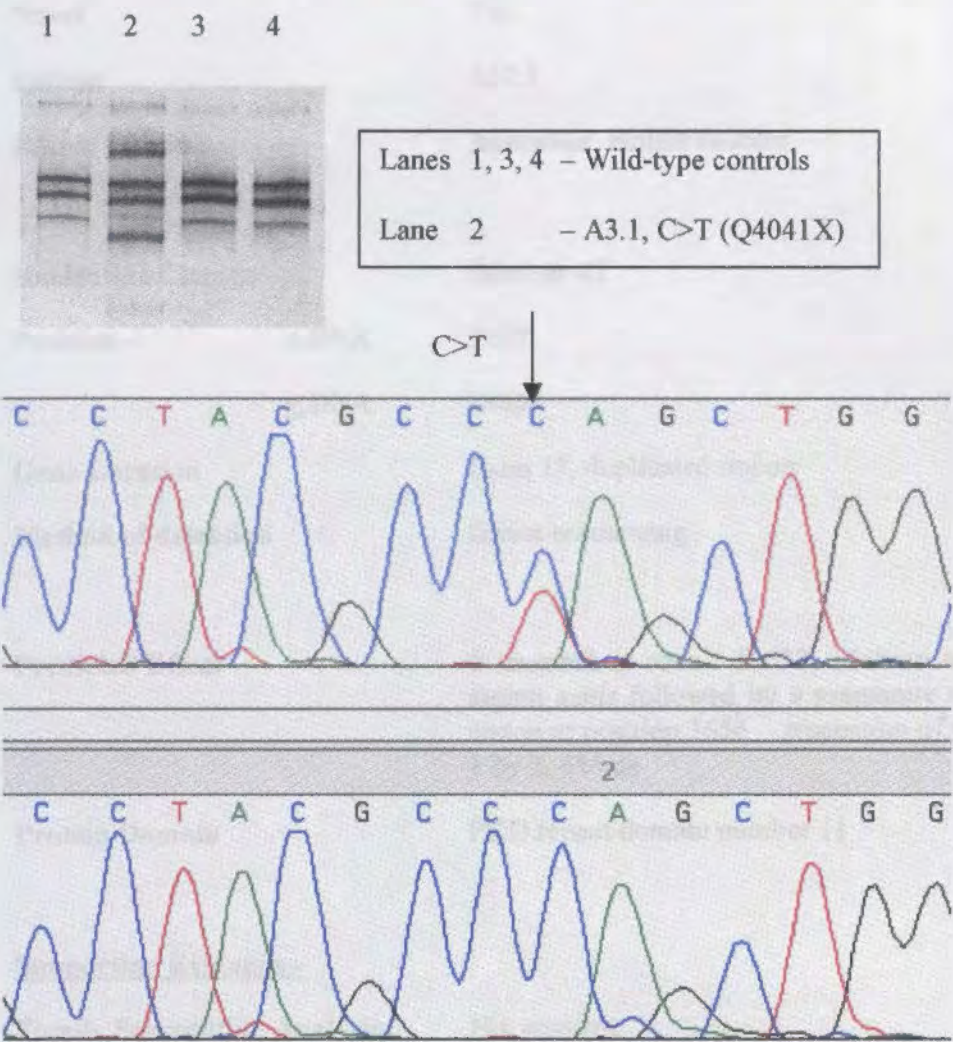


← 292bp

← 159bp

← 133bp

SSCP & Sequencing gels



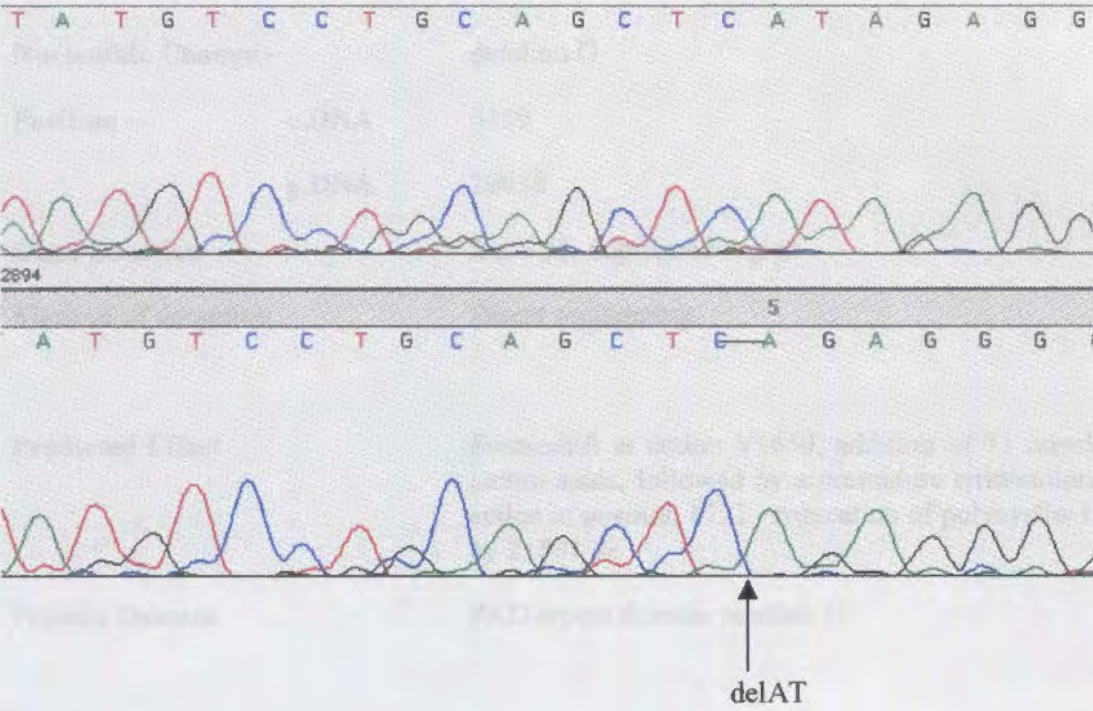
**Mutation – c.5105delAT**

<b>Novel</b>		Yes
<b>Patient</b>		A52.1
<b>Ethnic Origin</b>		Australian, British descent
<b>Nucleotide Change</b>		deletion AT
<b>Position –</b>	<b>c.DNA</b>	5105
	<b>g.DNA</b>	29004
<b>Gene Location</b>		Exon 15, duplicated region
<b>Method of detection</b>		Direct sequencing
<b>Predicted Effect</b>		Frameshift at codon T1632, addition of 24 novel amino acids followed by a premature termination codon at position 1656 - truncation of polycystin-1 by 2, 646 aa
<b>Protein Domain</b>		PKD repeat domain number 11
<b>Supporting Evidence -</b>		
<b>Family Segregation Analysis</b>		Not possible
<b>Confirmatory assay</b>		PCR repeated and products cloned into Topo-TA PCR cloning vector and sequenced



**Electrophoretogram**

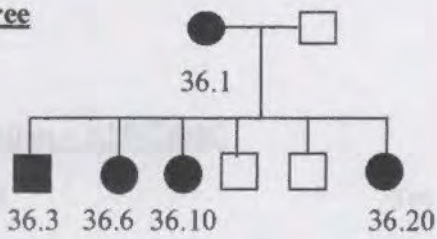
Upper row wild-type cloned product; lower row delAT cloned product



**Mutation – c.5159delG**

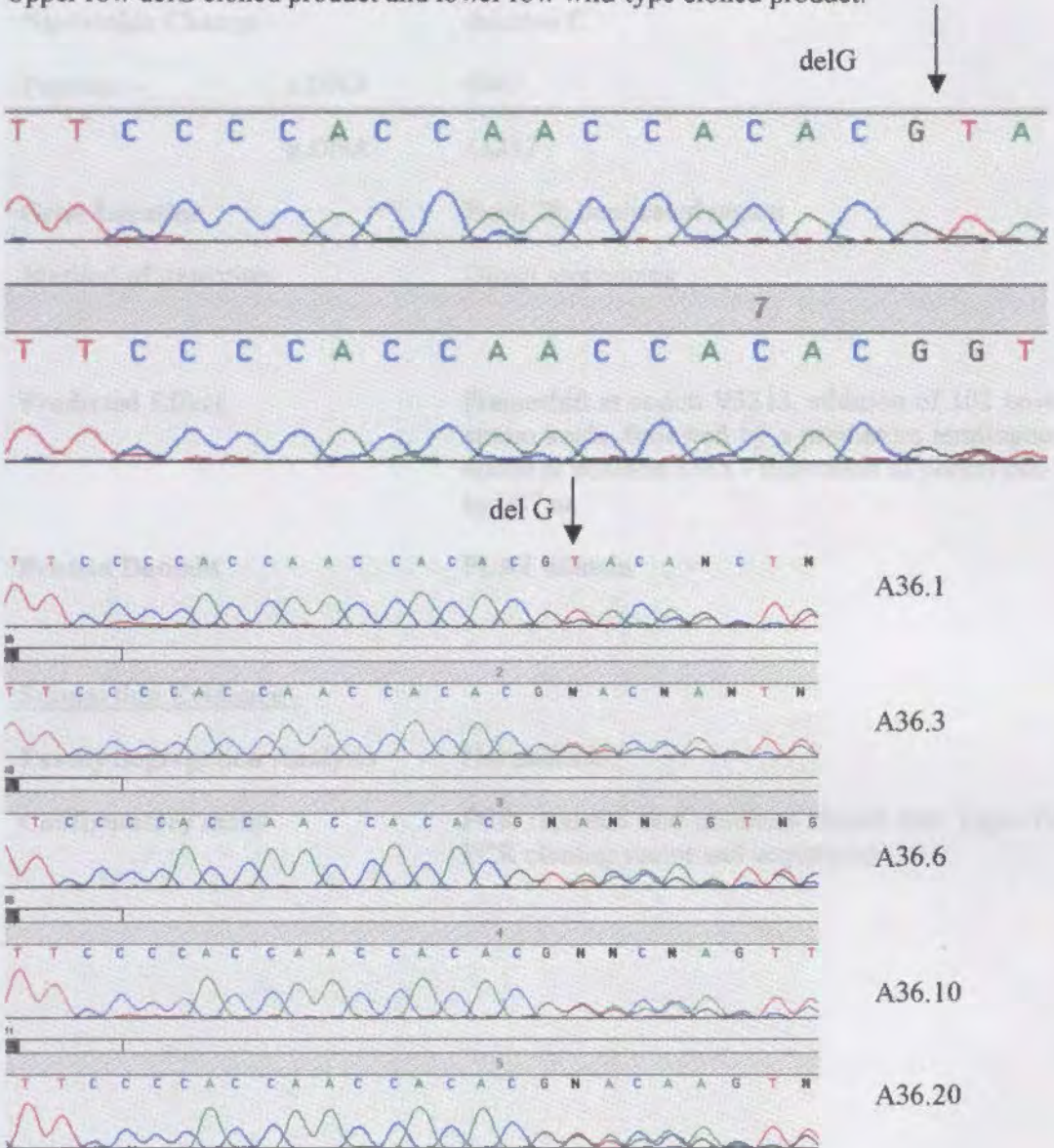
<b>Novel</b>	Yes
<b>Patient</b>	A36.1
<b>Ethnic Origin</b>	Australian, British descent
<b>Nucleotide Change -</b>	deletion G
<b>Position – c.DNA</b>	5159
<b>g.DNA</b>	29058
<b>Gene Location</b>	Exon 15, duplicated region
<b>Method of detection</b>	Direct sequencing
<b>Predicted Effect</b>	Frameshift at codon V1650, addition of 71 novel amino acids, followed by a premature termination codon at position 1721 - truncation of polycystin-1 by 2, 581 aa
<b>Protein Domain</b>	PKD repeat domain number 11
<b>Supporting Evidence -</b>	
<b>Family Segregation Analysis</b>	Segregates in 5 affected family members (A36.1, A36.3, A36.6, A36.10, A36.20). No unaffected members available for analysis
<b>Confirmatory assay</b>	PCR repeated and products cloned into Topo-TA PCR cloning vector and sequenced

## Pedigree



## Electrophoretogram

Upper row delG cloned product and lower row wild-type cloned product.

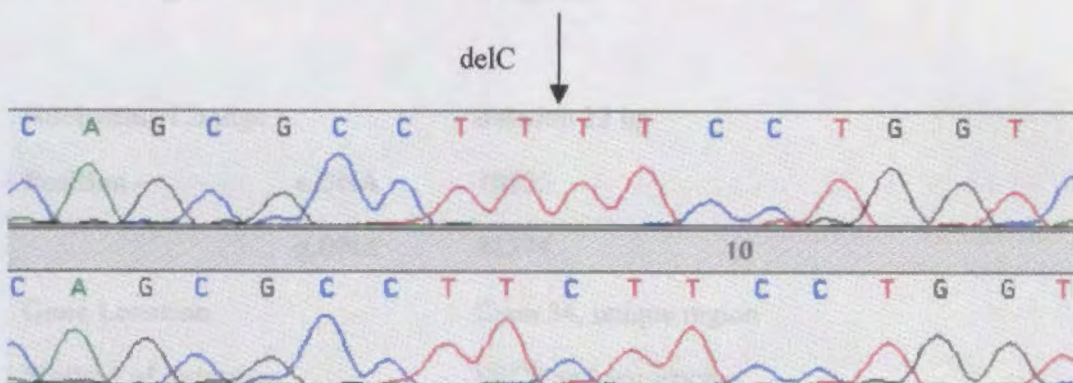




<b>Mutation – c.9847delC</b>		
<b>Novel</b>		Yes
<b>Patient</b>		A10.1
<b>Ethnic Origin</b>		Australian, British descent
<b>Nucleotide Change -</b>		
		deletion C
<b>Position –</b>	<b>c.DNA</b>	9847
	<b>g.DNA</b>	41352
<b>Gene Location</b>		Exon 28, duplicated region
<b>Method of detection</b>		Direct sequencing
<b>Predicted Effect</b>		Frameshift at codon V3213, addition of 102 novel amino acids, followed by a premature termination codon at position 3315 - truncation of polycystin-1 by 987 aa
<b>Protein Domain</b>		PLAT domain
<b>Supporting Evidence -</b>		
<b>Family Segregation Analysis</b>		Not possible
<b>Confirmatory assay</b>		PCR repeated and products cloned into Topo-TA PCR cloning vector and sequenced

# Electrophoretogram

Upper row delC cloned product; lower row wild-type cloned product



Sequenced 1 kb

Forward of vector 1475, addition of 41 nt of vector  
tail, followed by a primary termination signal at  
position 3317 - insertion of poly(GA) by T42 at

Frames Deleted

deletion between 1<sup>st</sup> & 4<sup>th</sup> T of sequence

Sequenced 1 kb

Sequenced 1 kb

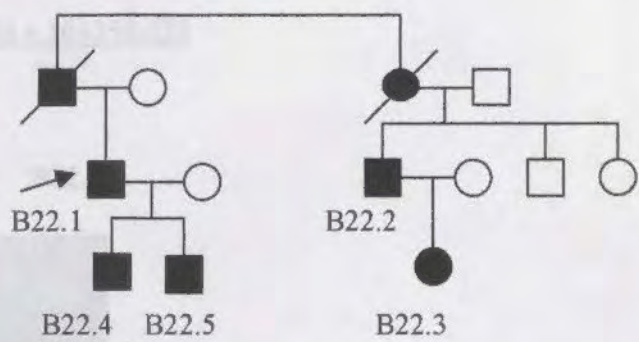
Sequencing of 7 nt of vector (1475, 1476, 1477, 1478, 1479, 1480, 1481) - the sequence  
sequence of the vector

Sequenced 1 kb

Sequencing of 7 nt of vector (1475, 1476, 1477, 1478, 1479, 1480, 1481) - the sequence  
sequence of the vector

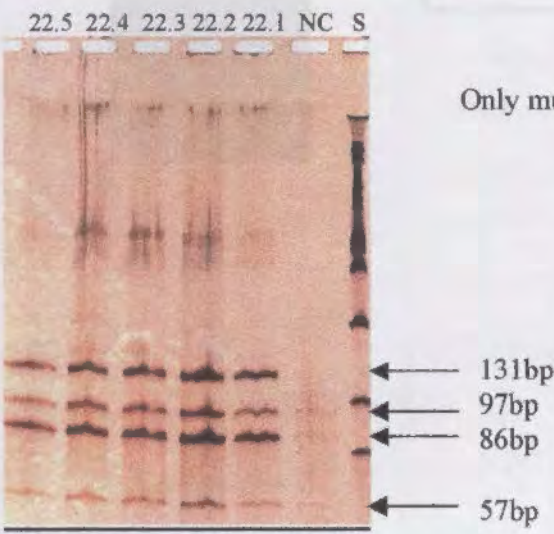
**Mutation – c.10635del23****Novel** Yes**Patient** B22.1**Ethnic Origin** Bulgarian**Nucleotide Change** deletion 23 bp**Position – c.DNA** 10635**g.DNA** 44374**Gene Location** Exon 34, unique region**Method of detection** SSCP and sequencing**Predicted Effect** Frameshift at codon 3475, addition of 42 novel amino acids, followed by a premature termination codon at position 3517 - truncation of polycystin-1 by 785 aa**Protein Domain** Helix between 5<sup>th</sup> & 6<sup>th</sup> TM domains**Supporting Evidence -****Family Segregation Analysis** Segregates in 5 affected family members (B22.1, B22.2, B22.3, B22.4, B22.5). No unaffected members available for analysis**Confirmatory assay** Ava II restriction enzyme assay (Restriction site abolished - 131 + 86 bp in mutant allele and 97, 57 + 86 bp in wild-type)

**Pedigree**



**AvaII Restriction Digest Assay**

Family B22 (22.5-22.1), normal control (NC) Size standard (S) (100bp ladder Gibco BRL)



Only mutant allele produces band of 131bp.



SSCP of B22.1 in centre lane with normal controls in surrounding lanes

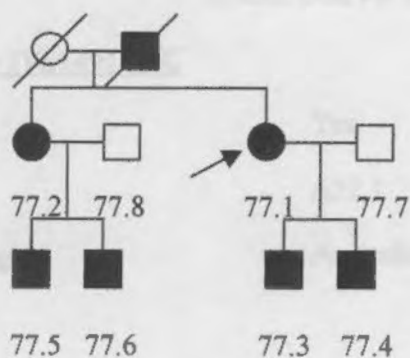
Normal B22.1



Deletion in sequence begins  
above the arrow

**Mutation – c.11587delG****Novel** Yes**Patient** B77.1**Ethnic Origin** Bulgarian**Nucleotide Change** deletion G**Position – c.DNA** 11587**g.DNA** 49515**Gene Location** Exon 40, unique region**Method of detection** SSCP and sequencing**Predicted Effect** Frameshift at codon 3792, addition of 32 novel amino acids, followed by premature termination codon at position 3824- truncation of polycystin-1 by 478 aa**Protein Domain** Loop between 6<sup>th</sup> and 7<sup>th</sup> TM domains**Supporting Evidence -****Family Segregation Analysis** Segregates in 6 affected family members (B77.1, B77.2, B77.3, B77.4, B77.5, B77.6), not found in the 2 unaffected spouses (B77.7, B77.8).**Confirmatory assay** SSCP and sequencing (no restriction enzyme assay possible)

## Pedigree



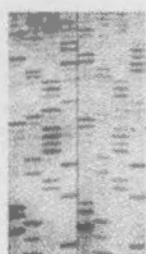
## SSCP

77.1 77.2 77.3 77.4 77.5 77.6 77.7 77.8



## Sequencing change

N M N: normal M: Mutant



delG

## **SPLICE-SITE VARIATION**

### **Mutation – IVS14+1G>C**

<b>Novel</b>	Yes
<b>Patient</b>	A22.1
<b>Ethnic Origin</b>	Australian, Spanish descent
<b>Nucleotide Change -</b>	G>C transversion
<b>Position –</b>	<b>c.DNA</b> 3506+1
	<b>g.DNA</b> 26942
<b>Gene Location</b>	IVS 14, duplicated region
<b>Method of detection</b>	Direct sequencing

**Predicted Effect** Canonical GT donor splice site abolished. Exact effect has not been confirmed, as RNA was not available. However, the most likely effect is inactivation of the splice site leading to translation into intron 14 for a further 14 aa before truncation at g.26953. Another possibility is the activation of a cryptic splice site 16 nucleotides beyond the canonical GT splice site at position g.26957-8; this would lead to the addition of 5 novel aa to polycystin-1.

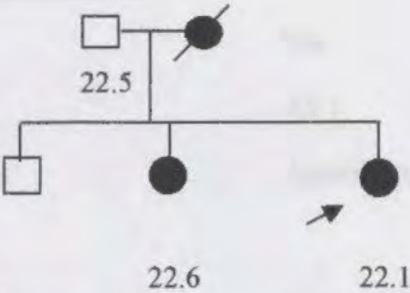
**Protein Domain** PKD repeat domain

### **Supporting Evidence -**

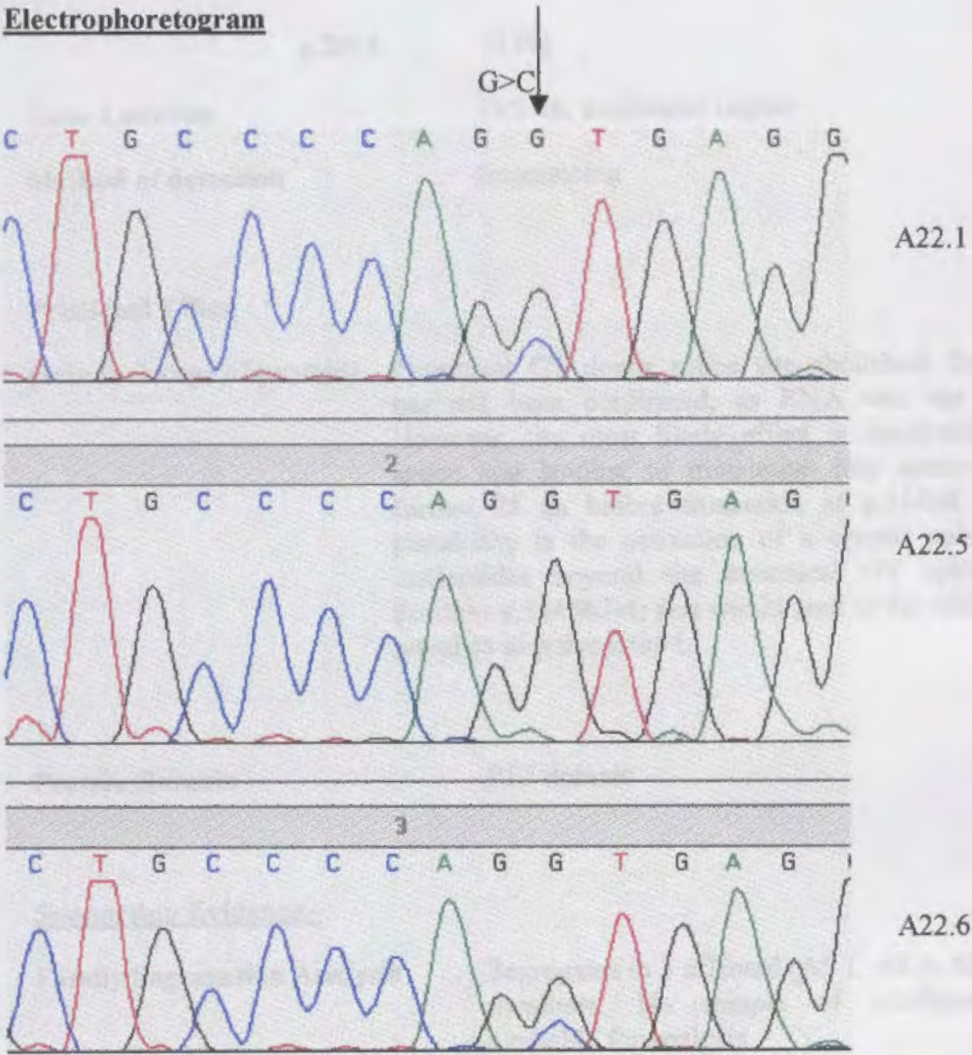
<b>Family Segregation Analysis</b>	Segregates in 2 affected (A22.1, A22.6), not found in 1 unaffected (A22.5) family members
<b>Confirmatory assay</b>	Sequencing (possible enzyme assay Alu I, restriction site created)



**Pedigree**



**Electrophoretogram**



**Mutation – IVS16+1G>T**

<b>Novel</b>	Yes
<b>Patient</b>	A5.1
<b>Ethnic Origin</b>	Australian, British descent

<b>Nucleotide Change</b>	G>T transversion
<b>Position –</b>	
<b>c.DNA</b>	7276+1
<b>g.DNA</b>	31394
<b>Gene Location</b>	IVS 16, duplicated region
<b>Method of detection</b>	Sequencing

**Predicted Effect**

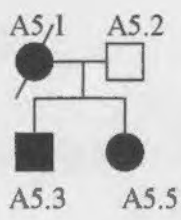
**(actual change of protein)** Canonical GT donor splice site abolished. Exact effect has not been confirmed, as R<sup>16</sup>A was not available. However, the most likely effect is inactivation of the splice site leading to translation into intron 16 for a further 25 aa before truncation at g.31468. Another possibility is the activation of a cryptic splice site 70 nucleotides beyond the canonical GT splice site at position g.314563-4; this would lead to the addition of 23 novel aa to polycystin-1.

<b>Protein Domain</b>	R <sup>16</sup> J domain
-----------------------	--------------------------

**Supporting Evidence -**

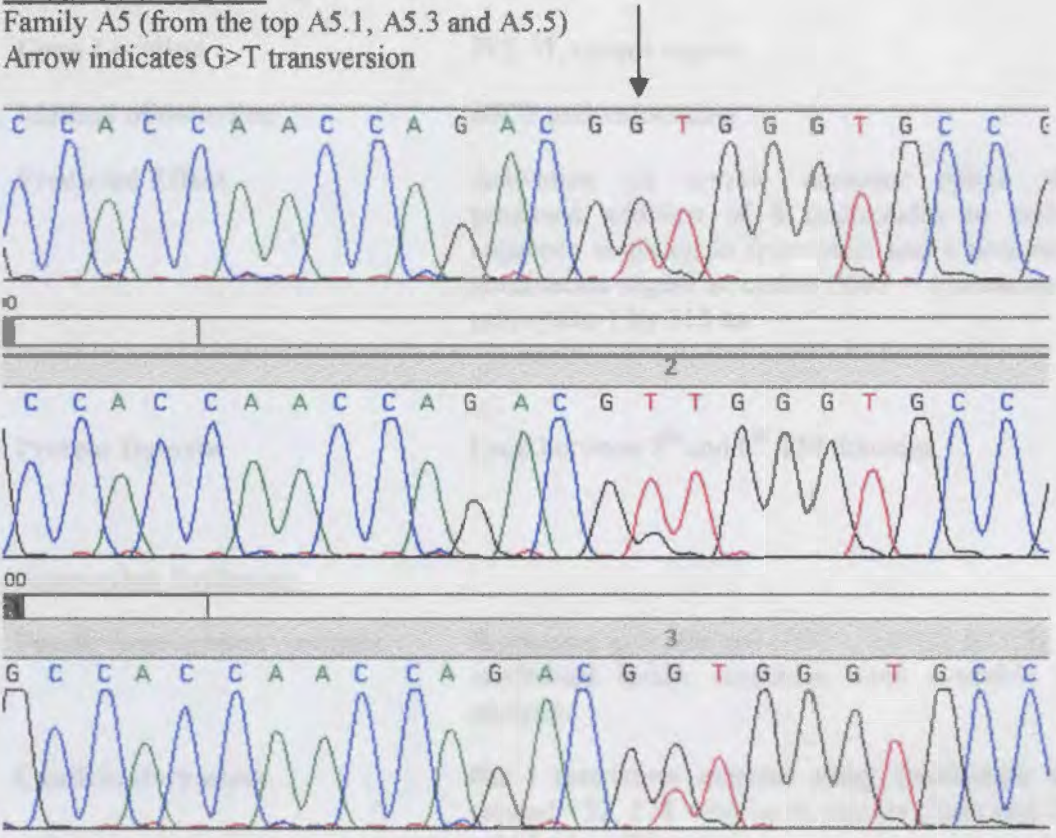
<b>Family Segregation Analysis</b>	Segregates in 3 affected (A5.1, A5.3, A5.5) family members. No sample of unaffected (A5.2) available for analysis.
<b>Confirmatory assay</b>	Sequencing (possible enzyme assay Mae II, restriction site created)

**Pedigree**  
Pedigree



**Electrophoretogram**

Family A5 (from the top A5.1, A5.3 and A5.5)  
Arrow indicates G>T transversion

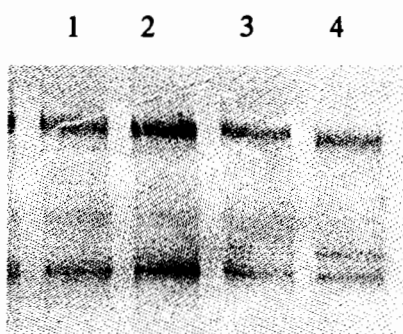
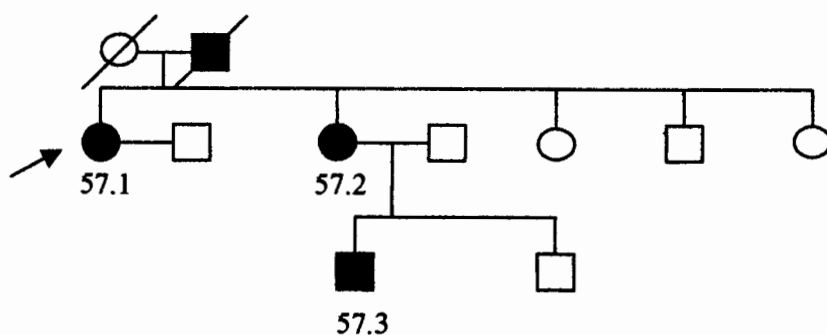


<b>Mutation – IVS37-10C&gt;A</b>		
<b>Novel</b>		Yes
<b>Patient</b>		B57.1
<b>Ethnic Origin</b>		Bulgarian
<b>Nucleotide Change</b>		C>A transversion
<b>Position –</b>	<b>c.DNA</b>	11225-10
	<b>g.DNA</b>	48491
<b>Gene Location</b>		IVS 37, unique region
<b>Method of detection</b>		SSCP and sequencing
<b>Predicted Effect</b>		Activation of cryptic acceptor splice site: predicted addition of 8 nucleotides to coding sequence resulting in frameshift and a premature termination signal at codon 3685 – truncation of polycystin-1 by 518 aa
<b>Protein Domain</b>		Loop between 5 <sup>th</sup> and 6 <sup>th</sup> TM domains
<b>Supporting Evidence -</b>		
<b>Family Segregation Analysis</b>		Segregates in 3 affected (B57.1, B57.2, B57.3), no unaffected family members were available for analysis.
<b>Confirmatory assay</b>		Bsr I restriction enzyme assay (restriction site created - 52, 174 + 46 bp in mutant allele and 226 + 46 bp in wild-type)

<b>Wild Type</b>	48481 agggcccccc cgtccacc [ag] AGC CTC CTG.....
	S    L    L
<b>Mutant</b>	48481 agggcccccc [ag] TCC ACC AGA GCC TCC TGG.....
	S    T    R    A    S    W

**Figure V.3** Proposed effect of splice site mutation g.48491C>A

### Pedigree

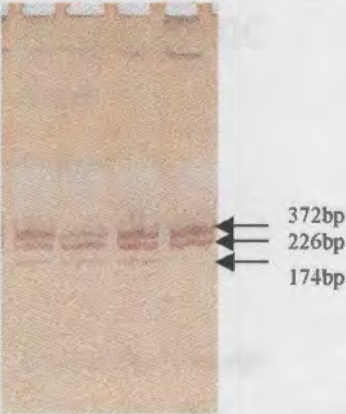


**SSCP Gel: PCR fragment IVS37-IVS 38**

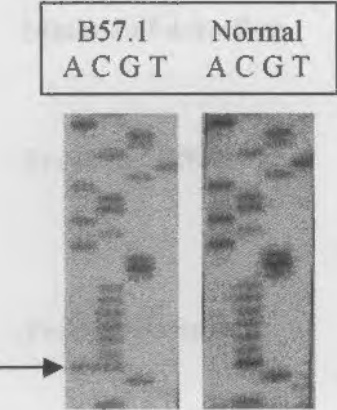
Lanes 1-3    Normal controls  
Lane 4        B57.1

**BsrI Restriction Digest**

57.1 57.2 57.3 NC      Family B57 with normal control (NC)



**Sequencing Gel**



Arrow indicates C>A mutation

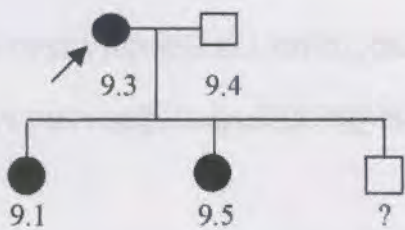
## MISSENSE VARIATION

### **Mutation – G381C**

<b>Novel</b>	Yes
<b>Patient</b>	A9.1
<b>Ethnic Origin</b>	Australian, British descent
<b>Nucleotide Change</b>	G>T transversion
<b>Position –</b>	
c.DNA	1352
g.DNA	21430
<b>Gene Location</b>	Exon 5, duplicated region
<b>Method of detection</b>	Sequencing
<b>Predicted Effect</b>	Non-polar glycine (conserved between mouse and human) is replaced by uncharged polar cysteine with a sulphhydryl group. Loop & helix changes in protein topology predicted by PHDsec (Fig V.4).
<b>Protein Domain</b>	Between 1 <sup>st</sup> PKD domain and C-type lectin domain
<b>Supporting Evidence -</b>	
<b>Family Segregation Analysis</b>	Segregates in 3 affected (A9.1, A9.3, A9.5), not found in 1 unaffected (A9.4) family members.
<b>Confirmatory assay</b>	Sequencing (possible Sac II restriction site abolished)

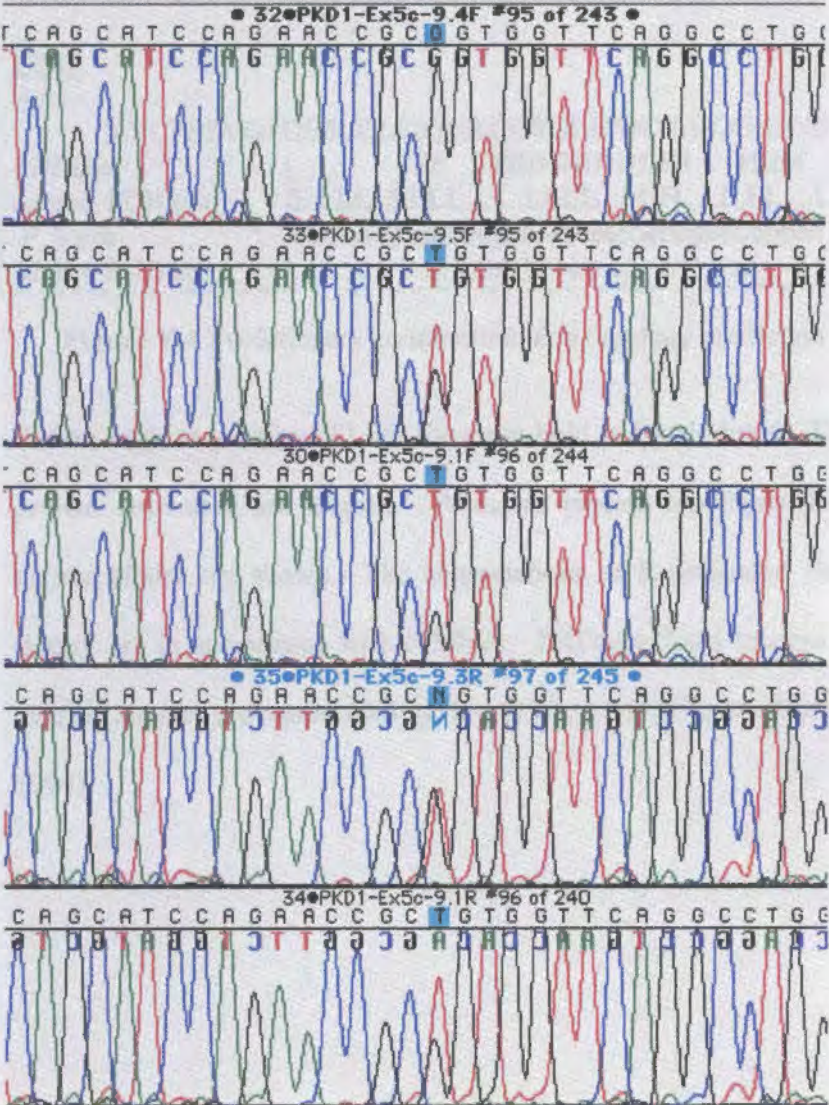


Pedigree



Electrophoretogram

G>T ↓



A9.4

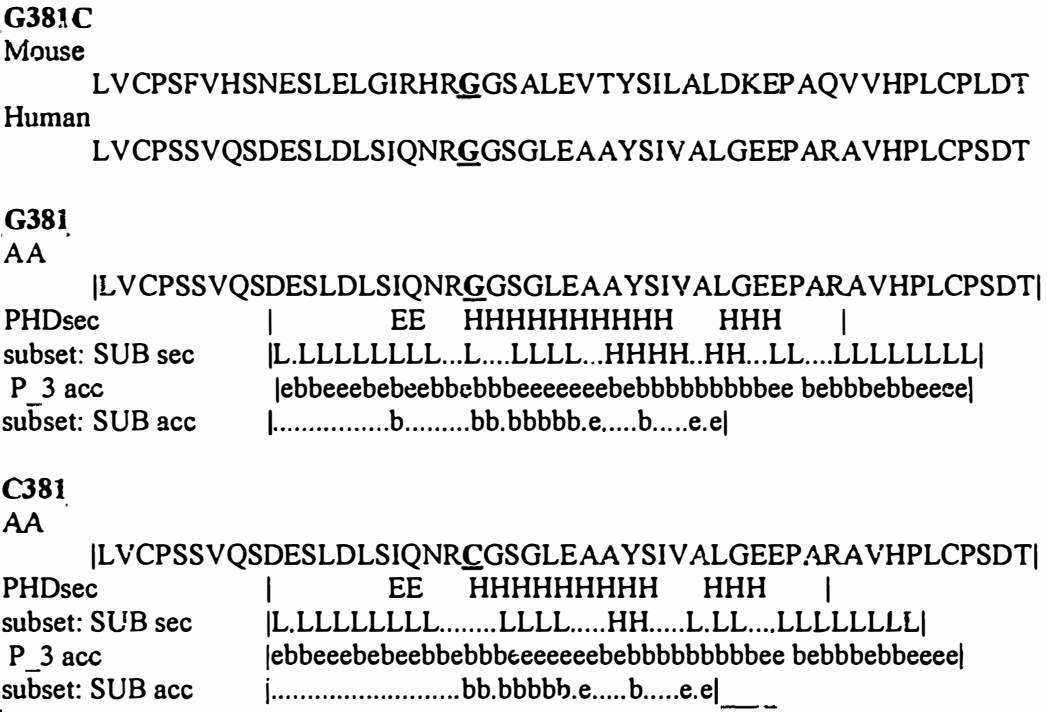
A9.5

A9.1F  
(forward  
reaction)

A9.3

A9.1R  
(reverse  
reaction)





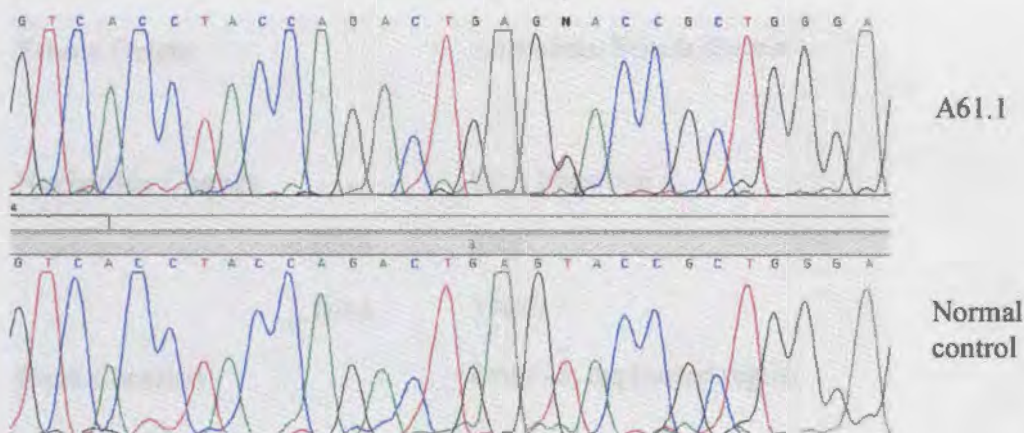
**Figure V.4** Evolutionary conservation and topology prediction for mutation G381C.

Amino acids at position 381 are shown in bold and underlined. The human and mouse protein sequences are aligned. Predicted protein topologies for the wild-type and variant alleles are shown. The abbreviations of E (extended sheet), H (helix) and L (loop), are in accordance with PHDsec. PHD predicted relative solvent accessibility (acc) in 3 states are represented by: b = 0-9%, i = 9-36%, e = 36-100%. (Rost & Sander 1994).

**Mutation – Y2185D**

<b>Novel</b>		Yes
<b>Patient</b>		A61.1
<b>Ethnic Origin</b>		Australian, British descent
<b>Nucleotide Change</b>		T>G transversion
<b>Position –</b>	<b>c.DNA</b>	6764
	<b>g.DNA</b>	30662
<b>Gene Location</b>		Exon 15, duplicated region
<b>Method of detection</b>		Sequencing
<b>Predicted Effect</b>		Non-polar tryptophan (conserved between pufferfish, mouse and human) is replaced by acidic aspartic acid. Loop lost in protein topology predicted by PHDsec (fig V.5).
<b>Protein Domain</b>		PKD domain
<b>Supporting Evidence -</b>		
<b>Family Segregation Analysis</b>		Not possible
<b>Confirmatory assay</b>		Sequencing (possible Ava II restriction site created)

## T&gt;G



Mouse QRNYLEAHVDLRNCVSYQTEYRWEIYRTASCQRPGRMAQMVLPGVD  
 Human QRNYLEAHVDLRDCVTYQTEYRWEVYRTASCQRPGRPARVALPGVD  
 Fugu LVEASVDLKGCLRYGAQYLWQILSAPSCDNDPHFASGRVNGAT

## AA

|QRNYLEAHVDLRDCVTYQTEYRWEVYRTASCORPGRPARVALPGVD|

PHD sec		EE	EE	EE	
subset: SUB sec		LLLE.....LL.....LLLL.....LLL.LLLLLLLLLL.....LLL			
P_3 acc		e ebbbbbbebbbbeeeeeeebeee eeee e ee ebebbbbebbe			
subset: SUB acc		e...b.b.b.b...b...e...e.....b.b.....e			

## AA

|QRNYLEAHVDLRDCVTYQTEDRWEVYRTASCQRPGRPARVALPGVD|

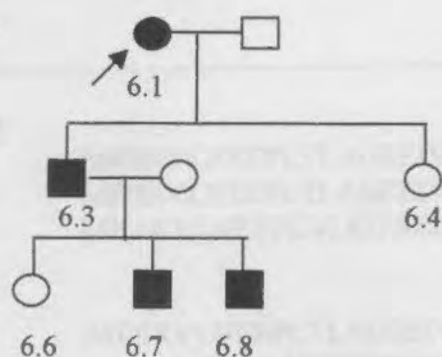
```
PHD sec      | EEEE EEE  EE  HHHHEEEEEEE E   EEEE
subset: SUB sec | LL..... LLLL..E..LLL|
P 3 acc       | eee beb bebeebbb ee bebb ebebe ee ebebbbbeeee|
```

**Figure V.5** Evolutionary conservation and topology prediction for mutation Y2185D.

**Mutation – G2785D**

<b>Novel</b>	Yes
<b>Patient</b>	A6.1
<b>Ethnic Origin</b>	Australian, British descent
<b>Nucleotide Change</b>	G>A transition
<b>Position – c.DNA</b>	8565
<b>g.DNA</b>	37893
<b>Gene Location</b>	Exon 23, duplicated region
<b>Method of detection</b>	Sequencing
<b>Predicted Effect</b>	Non-polar glycine (conserved between pufferfish, mouse and human) is replaced by acidic aspartic acid. Slight change, loop shortened, in protein topology predicted by PHDsec (fig V.6).
<b>Protein Domain</b>	Upstream of REJ domain in extracellular region prior to TM domains
<b>Supporting Evidence -</b>	
<b>Family Segregation Analysis</b>	Segregates in 4 affected (A6.1, A6.3, A6.8, A6.7), not found in 2 unaffected (A6.4, A6.6) family members.
<b>Confirmatory assay</b>	Sequencing (restriction assay not possible)

# Pedigree



# Electrophoretogram

10PKD1-EX23AF-6.1 #191 of 304  
T C G T G G C C C A G G N C A A G C G C T C G  
T C G T G G C C C A G G N C A A G C G C T C G

A6.1

16PKD1-ex23af-6.3 #172 of 285  
T C G T G G C C C A G G N C A A G C G C T C G  
T C G T G G C C C A G G N C A A G C G C T C G

A6.3

18PKD1-ex23af-6.6 #152 of 265  
T C G T G G C C C A G G G C A A G C G C T C G  
T C G T G G C C C A G G G C A A G C G C T C G

A6.6

20PKD1-ex23af-6.8 #150 of 263  
T C G T G G C C C A G G N C A A G C G C T C G  
T C G T G G C C C A G G N C A A G C G C T C G

A6.8

19PKD1-ex23af-6.7 #149 of 263  
T C G T G G C C C A G G N C A A G C G C T C G  
T C G T G G C C C A G G N C A A G C G C T C G

A6.7

17PKD1-ex23af-6.4 #143 of 257  
T C G T G G C C C A G G G C A A G C G C T C G  
T C G T G G C C C A G G G C A A G C G C T C G

A6.4

**G2785D**

Mouse MRSRVLN~~E~~EPLTLAGEEIV~~A~~L~~G~~KRSDPLSLLCY  
 Human MRSRVLN~~E~~EPLTLAGEEIV~~A~~Q~~G~~KRSDPRSLLCY  
 Fugu MHARVLN~~E~~EPLVLKGAEIAAT~~G~~KLADPQSLLCY

**G2785**

AA |MRSRVLN~~E~~EPLTLAGEEIV~~A~~Q~~G~~KRSDPRSLLCY|  
 PHD sec | EE HHHHHHHHHHHH EEEE |  
 subset: SUB sec |LL...LLL...HHHHHHHHH.LLLL.EEEL|  
 P\_3 acc |eeebbeebebbbbbbeebbebeeeeeebebbbe|  
 subset: SUB acc |.....e.....b....e.....b..|

**D2785**

AA |MRSRVLN~~E~~EPLTLAGEEIV~~A~~Q~~D~~KRSDPRSLLCY|  
 PHD sec | EE HHHHHHHHHHHH EEEE |  
 subset: SUB sec |LLL...LLL...HHHHHHHHH.LLLL.EEEL|  
 P\_3 acc |eeebbeebebbbbbbeebbebeeeeeebebbbe|  
 subset: SUB acc |.....e.....b....e.....b..|

**Figure V.6.** Evolutionary conservation and topology prediction for mutation G2785D.

**Mutation – L4046F**

<b>Novel</b>	Yes
<b>Patient</b>	A54.1
<b>Ethnic Origin</b>	Australian, British descent

**Nucleotide Change** C>A transversion

**Position – c.DNA** 12347

**g.DNA** 51003

**Gene Location** Exon 45, unique region

**Method of detection** Sequencing

**Predicted Effect** Non-polar leucine (conserved between mouse and human) is replaced by aromatic hydrocarbon – containing phenylalanine. Changes in protein topology predicted by PHDsec

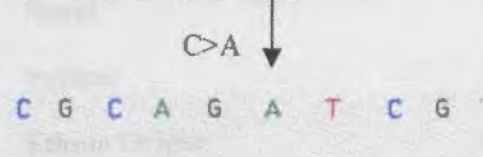
**Protein Domain** 9<sup>th</sup> TM domain

**Supporting Evidence -**

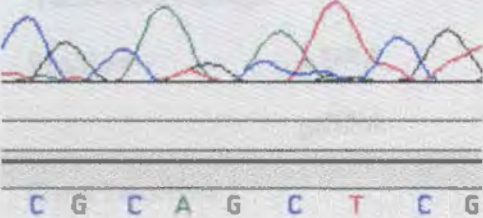
**Family Segregation Analysis** Not possible

**Confirmatory assay** Sequencing (possible Bbv I restriction site abolished).

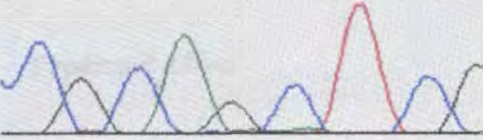
**Electrophoretogram L4046F**



A54.1



Normal control



**L4046F**

Mouse      AYAQMAILLISSGAD  
Human    AYAQL AILLVSSCVD

**4046L**

AA            |LGLVVLGVAYAQLAILLVSSCVDLSWSVAQALLVLCPGTG|  
PHD sec        | EE HHHHHHHHHHHHHHH HHHHHHHHHHHHHHHH |  
subset: SUB sec |LL.....HHHHHHHHHHHHH...HHHHHHHHHHHHHH..LLL|

**4046F**

AA            |LGLVVLGVAYAQLAILFVSSCVDLSWSVAQALLVLCPGTG|  
PHD sec        | EEEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH |  
subset: SUB sec |L.EEE..HHHHHHHHHHH.HH.HHHHHHHHHHHHHHHH.LLLL|

**Figure V.7.** Evolutionary conservation and topology prediction for mutation L4046F.



**Mutation – R4227P**

<b>Novel</b>	Yes
<b>Patient</b>	A38.1
<b>Ethnic Origin</b>	Australian, British descent

<b>Nucleotide Change</b>	G>C transversion
--------------------------	------------------

<b>Position –</b>	<b>c.DNA</b>	12891
-------------------	--------------	-------

	<b>g.DNA</b>	51637
--	--------------	-------

<b>Gene Location</b>	Exon 46, unique region
----------------------	------------------------

<b>Method of detection</b>	Sequencing
----------------------------	------------

<b>Predicted Effect</b>	Non-polar arginine (conserved between mouse and human) is replaced by charged basic proline. Changes in protein topology predicted by PHDsec
-------------------------	--

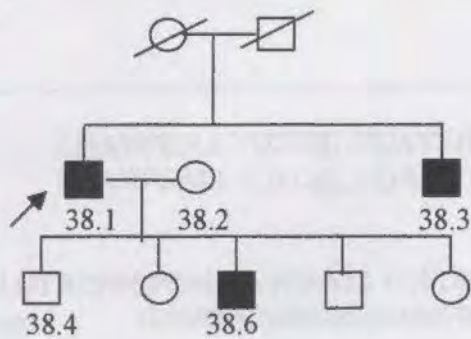
<b>Protein Domain</b>	Coiled-coil domain
-----------------------	--------------------

**Supporting Evidence -**

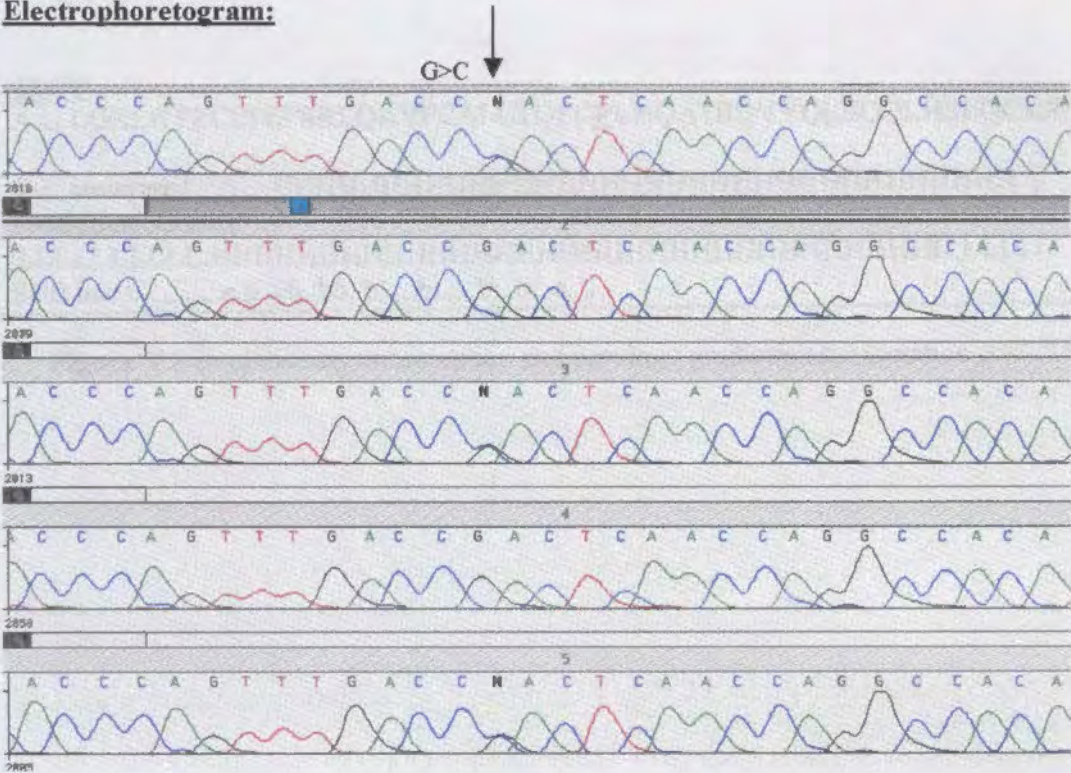
<b>Family Segregation Analysis</b>	Segregates in 3 affected (A38.1, A38.3, A38.6), not found in 2 unaffected (A38.2, A38.4) family members.
------------------------------------	--

<b>Confirmatory assay</b>	Sequencing (restriction assay not possible)
---------------------------	---

**Pedigree:**



**Electrophoretogram:**



From the top of the page these are A38.1, A38.2, A38.3, A38.4 and A38.6.

**R4227P**

Mouse LHA<sup>~</sup>VFESLLVQFDRLNQATEDVYQLEQQQLQSLRG  
Human LQAVFEALLTQFDRLNQATEDVYQLEQQQLHSLQG

**4227R**

AA: LGRLGTRCEPEPSRLQAVFEALLTQFDRLNQATEDVYQLEQQQLHSLQGRR|  
Sec. structure | HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH |  
Prediction Subsets  
|LLLLLLLLLLLLL..HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH.LLLL|  
SUB acc |.....bbb.bbb.....b...b..b.....|

**4227P**

AA: LGRLGTRCEPEPSRLQAVFEALLTQFDPLNQATEDVYQLEQQQLHSLQGRR|  
Sec. structure | HHH |  
Prediction Subsets  
|LLLLLLLLLLLLL.HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH.LLLL|  
SUB acc |.....e..b..bb..bb..b.....b.....b...b...e...|

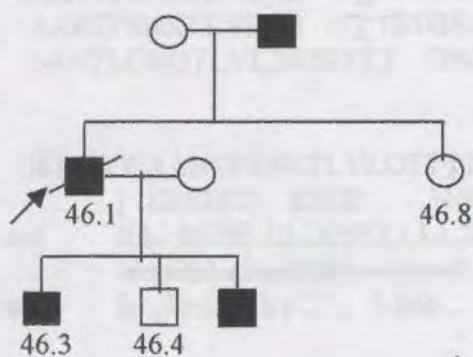
**Figure V.8** Evolutionary conservation and topology prediction for mutation R4227P.

## NON-TRUNCATING INSERTIONS & DELETIONS

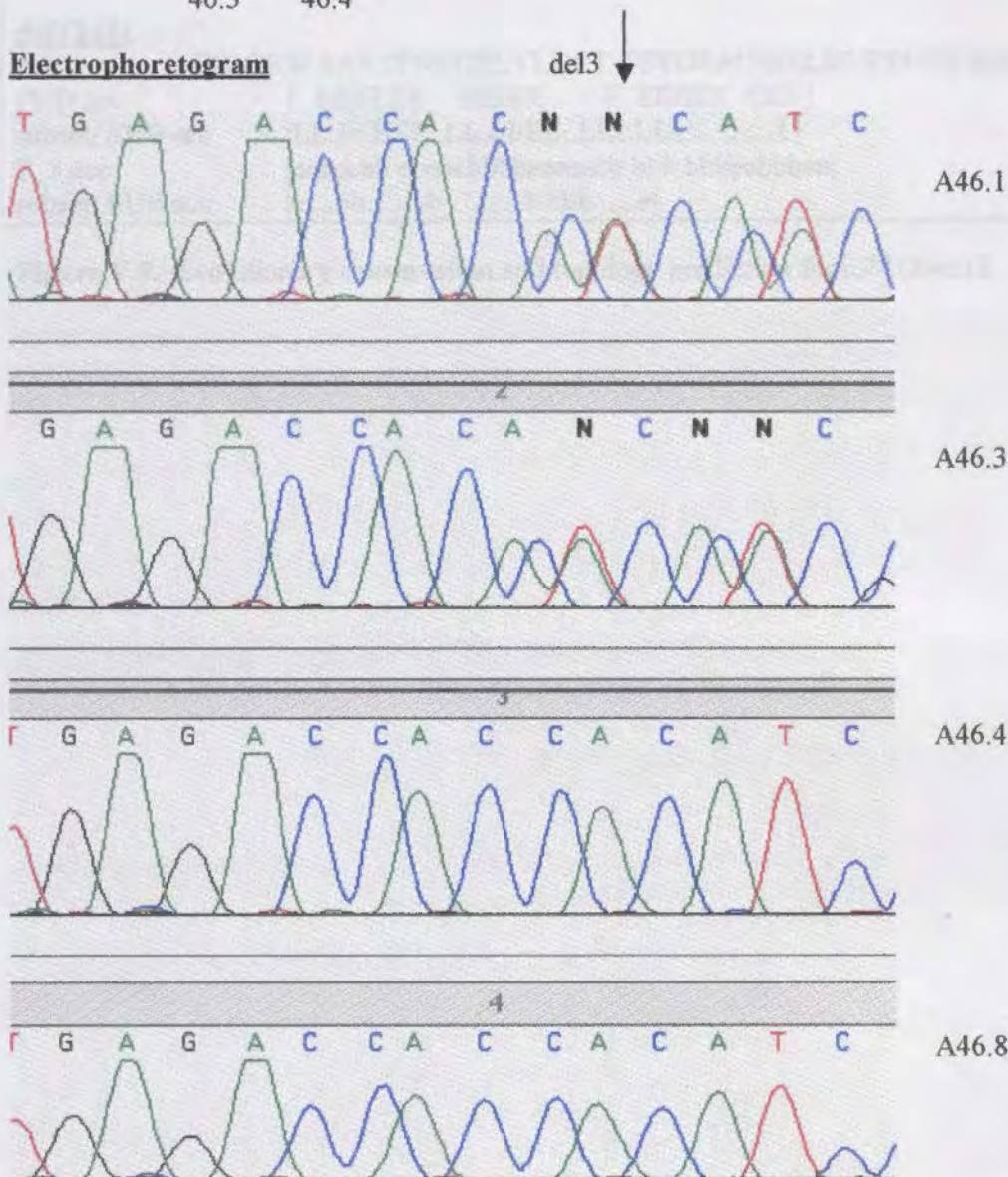
### **Mutation – c.7472del3**

<b>Novel</b>	Yes
<b>Patient</b>	A46.1
<b>Ethnic Origin</b>	Australian, British descent
<b>Nucleotide Change</b>	deletion ACC
<b>Position –</b>	<b>c.DNA</b> 7472
	<b>g.DNA</b> 32650
<b>Gene Location</b>	Exon 18, duplicated region
<b>Method of detection</b>	Sequencing
<b>Predicted Effect</b>	Deletion of threonine residue at codon 2421 (conserved between pufferfish, mouse and human). Changes in protein topology predicted by PHDsec
<b>Protein Location</b>	REJ domain
<b>Supporting Evidence -</b>	
<b>Family Segregation Analysis</b>	Segregates in 2 affected (A46.1, A46.3), not found in 2 unaffected (A46.4, A46.8) family members.
<b>Confirmatory assay</b>	Sequencing (restriction assay not possible)

# **Pedigree**



# **Electrophoretogram**



<b>c.7472del3</b>	
Mouse	AARTFSNKTLVLNE TTTSTGSTGMNLVVR
Human	AARTFSNKTLVLDE TTTSTGSAGMRLVLR
<i>Fugu</i>	SAVTLQN <del>DTL</del> VLDSSSTTT GSGGMNLVLR
<b>Wild-type</b>	
AA	KRGRWAARTFSNKTLVLDET-TTSTGSAGMRLVLRRGVLRD
PHD sec	EEEEE EEEE HHHEEEEE EEE
subset: SUB sec	LL.EEEEE.LL.EEE.LLLL.LL.....L
P_3 acc	eeeebb ebeeebbb <del>eeeeee</del> eb bbb bbbeebbbe
subset: SUB acc	e....bb.....b.e.....b.bbb.....e
<b>delT2421</b>	
AA	KRGRWAARTFSNKTLVLDET-TSTGSAGMRLVLRRGVLRD
PHD sec	EEEEE EEEEE E EEEEE EEE
subset: SUB sec	LL.EEEEE.LL.EEEE.LLLLL.....L
P_3 acc	eeeebb ebeeebbb <del>eeeeee</del> eb bbb bbbeebbbe
subset: SUB acc	e....bb.....b.....b.bbb.....e

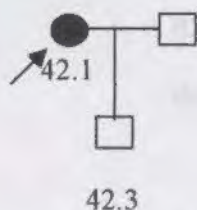
**Figure V.9.** Evolutionary conservation and topology prediction for c.7472del13.

**Mutation – c.9289del39**

Novel	Yes	
Patient	A42.1	
Ethnic Origin	Australian, British descent	
Change	39 bp deletion gac aga ggg gct gct gcc cct gga gga gac ctc gcc ccg	
Position –	c.DNA	9289
	g.DNA	39090
Gene Location	Exon 25, duplicated region	
Method of detection	Sequencing	
Predicted Effect	Deletion of aa 3027-3039: delTEGLLPLEETSPR (domain conserved between mouse and human). Changes in protein topology predicted by PHDsec	
Protein Location	Extracellular region prior to the 1st TM domain	
Supporting Evidence -		
Family Segregation Analysis	Detected in 1 affected (A42.1), absent in 1 unaffected (A42.3) family member. Linkage analysis in this family was performed on DNA from blood-spots. The yield was not sufficient for XL-PCR.	
Confirmatory assays	PCR repeated and products cloned into Topo-TA PCR cloning vector then sequenced. Agarose gel electrophoresis also showed both alleles.	

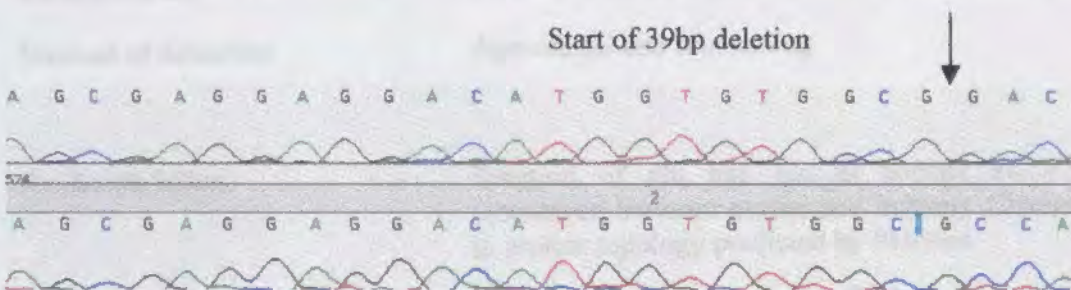


1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 2678, 2679, 26



### Electrophoretogram

AGGAGGACAT GGTGTGGCGG ACAGAGGGGC TGCTGCCCTT GGAGGAGACC TCGCCCGGCC AGGCCGCTCTG C  
AGGAGGACAT GGTGTGGCG- 120 150 180 210 240 270 300 330 360 390 420 450 480 510 540 570 600 630 660 690 720 750 780 810 840 870 900 930 960 990 -C AGGCCGTTTG C



c.9289del39

Human GLYTSLCQYFSEEDMVWR**TEGLLPLEETSPRQ**AVCLTRHLTA  
 Mouse GLYTSLCQYFSEEMMMWR**TEGIVPLEETSPSQ**AVCLTRHLTA

### Wild-type c.9289

AA	GLYTSLCQYFSEEDMVWR-TEGLLPLEETSPR-QAVCLTRHLTA		
PHD sec	EE	HHHH	EEEE
subset: SUB sec	L.EE.....LL.LL.H...LLLLLLLLLLLLLLL.L.....L		
P_3 acc	ebbbbbb bbbbe eeeeeeebbbebeeebeeebeeebeee		
subset: SUB acc	e.b.bbb.bb.....e.....e.ee.....e		

c.9289del39

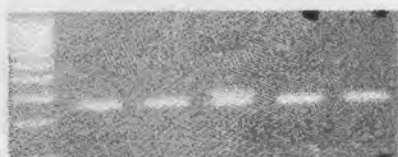
AA	GLYTSLCQYFSEEDMVWR--QAVCLTRHLTA
PHD sec	EEEHHHH HHHHHHHHHHHHHHHH
subset: SUB sec	L.EE.HH.....HHHHHHHHHHHHHHHLL
P_3 acc	ebbbbbbe bbee bbb bbbbbbbeebee
subset: SUB acc	e...bbb.....bbbbbb....e

**Figure V.10.** Evolutionary conservation and topology prediction for c.9289del39.



**Mutation – c.11669del6****Novel** Yes**Patient** B25.1**Ethnic Origin** Bulgarian**Nucleotide Change** 6 bp deletion CAGGAG**Position – c.DNA** 11669-11674**g.DNA** 49737-49742**Gene Location** Exon 41, unique region**Method of detection** Agarose gel and sequencing**Predicted Effect** Deletion of glu and gln at codons 3820-21 (conserved between mouse and human). Changes in protein topology predicted by PHDsec**Protein Domain** Extracellular region between 6<sup>th</sup>-7<sup>th</sup> TM domains**Supporting Evidence -****Family Segregation Analysis** Not possible**Confirmatory assay** Alu I restriction enzyme assay (restriction site abolished in mutant allele (resulting fragments are 214 and 135bp. Fragments are 172, 48, and 135bp in wild-type.

S N N B25.1 N N



# Agarose gel of Exon 41

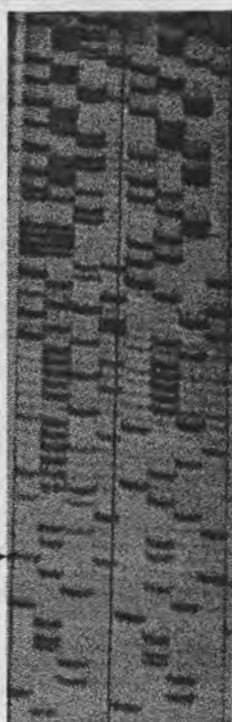
S- Size Standard

N- Normal Control

B25.1- c.11669del6 (Agarose gel appeared to have normal and larger allele;  
when sequenced this proved to in fact be a 6 bp deletion).

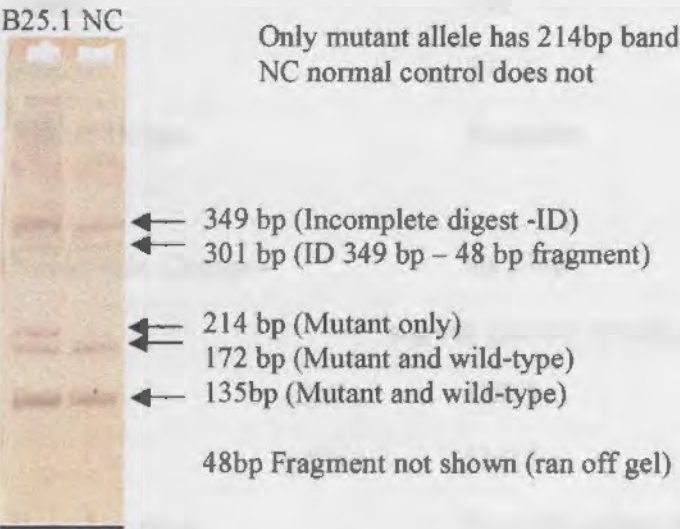
# Sequencing Gel c.11669del6

B25.1	Normal
ACGT	ACGT



Arrow indicates start of 6bp deletion

B25.1 AluI restriction assay



<b>c.11669del6</b>	
Human	DSGGYV <u>Q</u> ELGLSLEESRDRLRFLQLHNWL
Mouse	DSGGY I <u>Q</u> ELGLSLEESRARLGFLQLHNWL
<b>Wild-type c.11669</b>	
AA	AWSWGSCAVYDSGGYV <u>Q</u> ELGLSLEESRDRLRFLQLHNWLDN
sec. structure	EEE EEEEE HHHHHHHHHHHHHHHHH
prediction	
subset	LLLLL E LLL E HHHHHHHHH LL
<b>c.11669del6</b>	
AA	AWSWGSCAVYDSGGYVLGLSLEESRDRLRFLQLHNWLDN
sec. structure	EEEE EEEE HHHHHHHHHHHHHHHHH
prediction	
subset	LL L EEE LLL EEE HHHHHHHHHHHHH LL

Figure V.11. Evolutionary conservation and topology prediction for c.11669del6.

**Mutation – c.13069ins39**

Novel	Yes		
Patient	B41.1		
Ethnic Origin	Bulgarian		
Nucleotide Change -	ins 39bp		
	^acg cag gga cca ctt cgg gcc aag aac aag gtc cac ccc^		
Position –	c.DNA	13069	
	g.DNA	51815	
Gene Location	Exon 46, unique region		
Method of detection	SSCP and sequencing		
Predicted Effect	Insertion of 13 aa (TQGPLRAKNKVHP) between codons 4286-7 Changes in protein topology predicted by PHDsec		
Protein Domain	Cytoplasmic terminal region		
<u>Supporting Evidence -</u>			
Family Segregation Analysis	Segregates in 2 affected (B41.1, B41.5) and 2 pre-symptomatic (B41.4, B41.7), shown to be carriers by linkage analysis, absent in 3 unaffected (B41.2, B41.3, B41.6) family members.		
Confirmatory assay	Size-separation by agarose gel electrophoresis		

### Nucleotide Insertion

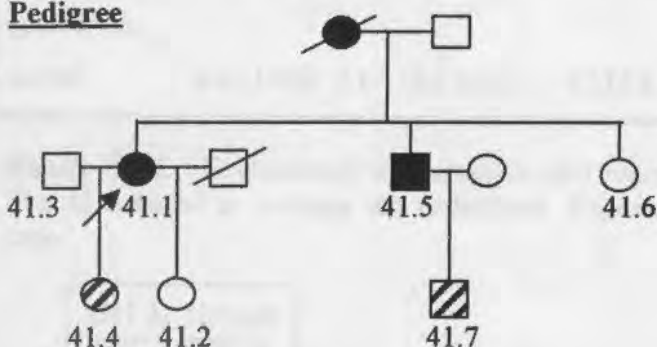
(4286) CCC ^acg cag gga cca ctt cgg gcc aag aac aag gtc cac ccc ^ AGC(4287)

### Additional Amino Acids

(4286) Pro ^Thr Gln Gly Pro Leu Arg Ala Lys Asn Lys Val His Pro^ Ser(4287).

Fig V.12. Nucleotide insertion and corresponding amino acids (c.13069ins39)

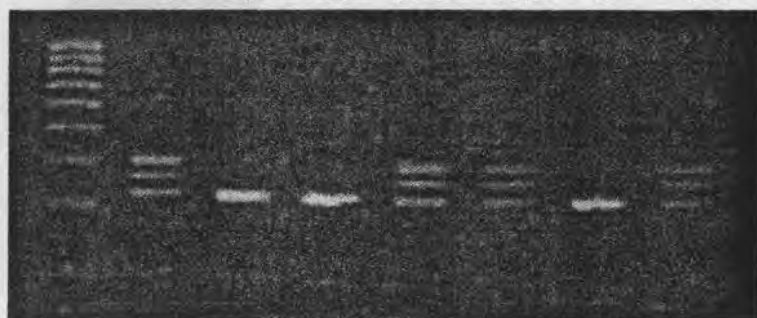
### Pedigree



Individuals B41.4 and B41.7 both display the insertion; both children are asymptomatic at this time.

### Family B41

S 41.1 41.2 41.3 41.4 41.5 41.6 41.7



### Agarose Gel

Lane 1:  
Marker (S)

Lanes 2-8:  
Family B41

(S) size standard – 100bp ladder

NB. Although 3 bands were seen on agarose gel in the affected individuals when this PCR product was sequenced only 2 alleles were detected, namely the normal wild type allele and the allele containing the 39 bp insertion as described. Future studies could include sequence analysis of individual bands to solve this anomaly.

# Polycystin-1, pos. 4268 – 4302

ex. 46, normal sequence

AA |ALPSRLARASRGVDLATGPSRTPLRAKNKVHPSST|

sec. structure | HHHHHHH |

prediction

subset |LL HHHHH LL LLLLLL LLLLLLL|

ex. 46, c.13069-13070ins39

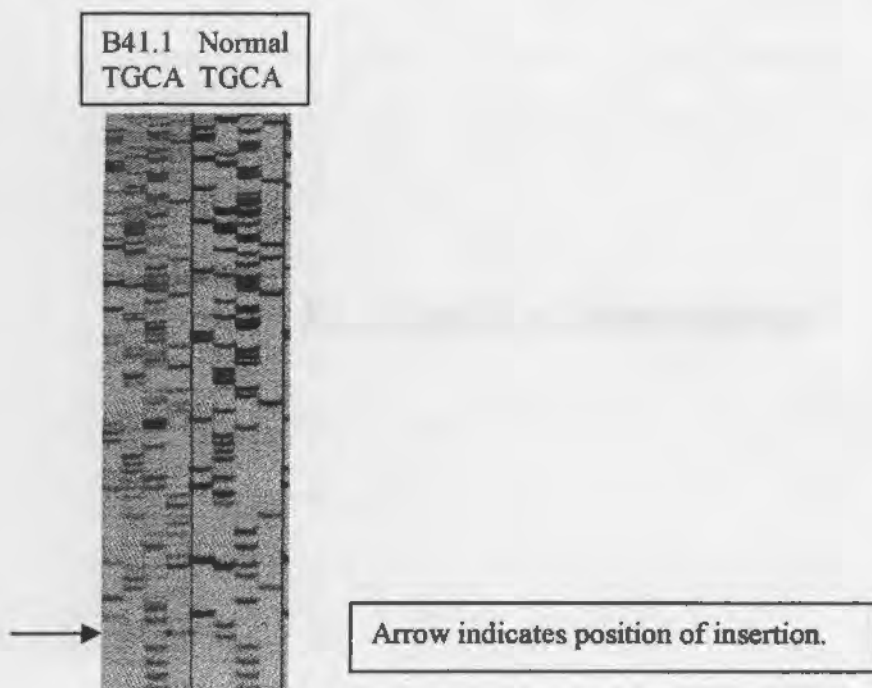
AA |ALPSRLARASRGVDLATGPTQRPLRAKNKVHPSRTPLRAKNKVHPSST|

sec. structure | HHHHHHH |

prediction

subset |LL H HH LL LLLLLL LLLLLLL LLLLLLL|

**Figure V.13.** Evolutionary conservation and topology prediction for c.13069ins39. The 13 inserted aa residues are underlined. Repeated amino acids are shown in bold case.



## **VI. Results – Polymorphisms**

## VI.1. Introduction

Alterations in the DNA sequence that have no apparent pathological significance and do not segregate with the disease phenotype in the affected families were classified as polymorphisms. Since the primary aim of this study was to identify the probable disease-causing mutations in *PKDI*, polymorphisms were detected as a “side product”, using the same analytical methodology as that used for the detection of pathogenic mutations (outlined in Chapter III.4).

When SSCP was used as a method for mutation screening, and several unrelated affected individuals produced identical aberrant banding patterns, they were assumed to share a common polymorphism. In such cases, sequencing analysis was performed on a small number of representative samples. After the sequencing identification of the variant causing the SSCP mobility shift, a confirmatory restriction assay was designed and performed on all DNA samples.

A total of 50 sequence variants in *PKDI* that did not segregate with the disease phenotype were detected in the families investigated (Tables VI.1 & VI.2). Nineteen of these variants have not been described previously.

Eight polymorphisms were intronic and the remaining 42 occurred in the coding sequences. This ratio is not likely to represent the true proportion of intronic to exonic polymorphisms



in *PKDI*, as it reflects the biased design of the study, where only 50-100 bp of flanking intronic sequence were included in the analysis of each exon.

All exonic polymorphisms were single nucleotide substitutions (SNPs). The overall length of the coding *PKDI* sequence analysed in this study was 12,163 bp, with a total of 42 SNPs detected. This allows an approximate estimate of 1 SNP occurring per 290 nt in this part of the genome.

Of the 42 exonic SNPs detected, 29 were silent, i.e. they did not result in an amino acid substitution. These variants, together with the intronic polymorphisms, are shown in Table VI.1.

#### VI.2. Missense Polymorphisms

A particularly interesting group of polymorphisms comprised 13 SNPs that resulted in missense variations in polycystin-1, yet did not segregate with the disease. These sequence variations are shown in Table VI.2. In general, missense changes present a challenge in mutation analysis. Their causative role in relation to the disease phenotype is difficult to establish, especially in cases such as ADPKD, where the role of the protein involved is complex and poorly understood, and *in vitro* functional assays are not available. In addition, even if unrelated to the disease phenotype, these variants may exert subtle functional effects and thus modify phenotypic severity – a hypothesis that is to be tested in future research.

**Table VI.1.** Silent mutations in the *PKD1* gene.

Location	Patients with change	Sequence Variation	Amino Acid Posn.	Type of Change	Previous Reference
Exon 2	A7.1	c.487G>A g.19963G>A	A92	Silent exonic	Rossetti <i>et al.</i> , (2001)
Exon 5	A36.1	c.1234C>T g.21312C>T	A341	Silent exonic	Thomas <i>et al.</i> , (1999)
Exon 5	A5.1 A9.1 A13.3 A15.1* A36.1 A42.1	c.1330T>C g.21408T>C	L373	Silent exonic	Rossetti <i>et al.</i> , (2001)
IVS 6	A7.1	g.22160G>A	-	Intronic	Novel
IVS 9	A9.1 A13.3 A36.1	g.23307del7 (VNTR)	-	Intronic	Novel
Exon 11	A5.1	c.2905A.C g.24952A>C g.	A898	Silent exonic	Thomas <i>et al.</i> , (1999)
Exon 11	A9.1 A13.3 A36.1	c.2911G>A g.24958G>A	P900	Silent exonic	Thomas <i>et al.</i> , (1999)
Exon 11	A9.1 A13.3 A36.1	c.2941C>T g.24988C>T	D910	Silent exonic	Thomas <i>et al.</i> , (1999)
Exon 13	A6.1 A29.1 A46.1	c.3274T>C g.26395T>C	G1021	Silent exonic	Thomas <i>et al.</i> , (1999)
Exon 13	A5.1	c.3319G>T g.26440G>T	T1036	Silent exonic	Novel

Exon 13	A7.1 A10.1 A52.1 A61.1	c.3322A>G g.26443A>G	L1037	Silent exonic	Thomas <i>et al.</i> , (1999)
Exon 15	A9.1 A46.1	c.3583C>T g.27483C>T	A1124	Silent exonic	Thomas <i>et al.</i> , (1999)
Exon 15	A9.1 A46.1	c.3586C>T g.27486C>T	S1125	Silent exonic	Thomas <i>et al.</i> , (1999)
Exon 15	A9.1 A13.3 A29.1 A36.1 A46.1	c.4876A>C g.28775A>C	A1555	Silent exonic	Watnick <i>et al.</i> , (1999)
Exon 15	A7.1 A10.1	c.4885G>A g.28784G>A	T1558	Silent exonic	Watnick <i>et al.</i> , (1999)
Exon 15	A9.1 A13.3 A29.1 A36.1 A46.1	c.5383C>T g.29280C>T	T1724	Silent exonic	Watnick <i>et al.</i> , (1999)
Exon 15	A9.1 A13.3 A36.1	c.5974G>A g.29872G>A	L1921	Silent exonic	Thomas <i>et al.</i> , (1999)
Exon 16	A13.3	c.7138C>T g.31255C>T	G2309	Silent exonic	Perrichot <i>et al.</i> , (2000b)
IVS 16	A42.1	g.31414 G>A	-	Intronic	Novel
Exon 17	A9.1 A13.3 A36.1 A46.1	c.7376T>C g.32427T>C	L2389	Silent exonic	Watnick <i>et al.</i> , (1999)
Exon 18	A13.3 A29.1 A36.1 A46.1	c.7652C>T g.32830C>T	L2481	Silent exonic	Perrichot <i>et al.</i> , (2000b)
Exon 21	A61.1	c.8098G>A g.33824G>A	A2629	Silent exonic	Novel

Exon 26	A6.1 A9.1 A29.1 A36.1 A46.1	c.9541T>C g.39466T>C	P3110	Silent exonic	Peral <i>et al.</i> , (1997)
IVS 31	A61.1	g.4 3740T>C	-	Intronic	Novel
Ex 35	B71.1  11/48	c.10743C>A g.47419C>A	A3511	Silent exonic	Novel
Ex 36	B62.4  8/48	c.10951C>T g.47705C>T	P3580	Silent exonic	Novel
Ex 36	A29.1, P113.1  2/48	c.10976C>T g.47705C>T	L3589	Silent exonic	Aguiari <i>et al.</i> , (2000)
IVS 38	A25.1, A34.1  3/48	c.IVS38+13G>A g.48653G>A	-	Intronic	Perrichot <i>et al.</i> , (1999)
Ex 40	B23.1  1/48	c.11554C>T g.49482C>T	D3781	Silent exonic	Novel
Ex 40	A4.1, A22.1  2/48	c.11584G>C g.49512G>C	S3791	Silent exonic	Peral <i>et al.</i> , (1996a)
IVS 41	B23.1**  1/48	c.IVS41+5ins3 g.49818insGGG	-	Intronic	Perrichot <i>et al.</i> , (2000a)
IVS 44	P107.1  1/50	c.IVS44+12C>T g.50931C>T	-	Intronic	Novel
IVS 44	P113.1, A29.1  2/50	c.IVS44+23insA g.50942insA	-	Intronic	Novel

Ex45	A4.1, B1.1, B78.1*	c.12484A>G g.51140A>G	A4091	Silent exonic	Aguiari <i>et al.</i> , (2000)
	21/48				
Ex 46(a)	B4.1 B78.1*	c.12838T>C g.51584T>C	4209P	Silent exonic	Peral <i>et al.</i> , (1997)
	17/48				
Ex 46(a)	B3.1	c.12871C>T g.51617C>T	4220A	Silent exonic	Novel
	1/48				
Ex 46(a)	A3.1	c.12872C>T g.51618C>T	4221L	Silent exonic	Novel
	1/48				

\*A15.1 homozygous for c.1330T>C, & \*B78.1 homozygous for c.12484A>G and c.12838T>C. All other individuals are heterozygotes for the polymorphisms listed.

\*\* IVS41+5ins3 previously reported as disease-causing by Perrichot *et al.*, (2000a) then as a rare neutral polymorphism by Koptides *et al.*, (2000).

**Table VI.2.** Missense variations in *PKDI* that did not segregate with the disease

Exon	Patients with change	Sequence Variation	Amino Acid change	Frequency (%) *	Previous Reference
Ex 11	A42.1	c.2427A ** g.24474G>A	Q739R	1/34 (2.9%)	Novel
Ex 14	A6.1, A29.1, A46.1	c.3486T>C g.26921T>C	M1092T	3/34 (8.8%)	Novel
Ex 15	A29.1, A46.1	c.4406T>C g.28306T>C	W1399R	2/34 (5.8%)	Watnick <i>et al.</i> , (1999)
Ex 15	A52.1	c.5157C>T g.29056C>T	T1649M	2/34 (5.8%)	Novel
Ex 21	A29.1, A46.1	c.8124A>G g.33849A>G	H2638R	2/34 (5.8%)	Watnick <i>et al.</i> , (1999)
Ex 23	A42.1, A46.1	c.8504C>T g.37844C>T	R2765C	2/34 (5.8%)	Novel
Ex 25	A6.1, A9.1, A29.1, A36.1, A46.1	c.9406GT>CC g.39207GT>CC	F3066L	5/34 (14.7%)	Peral <i>et al.</i> , (1997)
Ex 35	B80.1 P108.1	c.10737C>T g.47413C>T	T/M3509	2/96 (2.1%)	Peral <i>et al.</i> , (1997).
Ex 36	B1.1 A29.1	c.10986G>T g.47737G>T	S/I3592	2/96 (2.1%)	Novel
Ex 38	P109.1	c.11295G>A g.48571G>A	G/Q3695	1/96 (1.0%)	Novel
Ex 40	P101.1	c.11582T>A g.49510T>A	S/T3791	1/96 (1.0%)	Novel
Ex 44	22 people B78.1***	c.12341A>G g.50914A>G	I/V4044	23/100 (23.0%)	Rossetti <i>et al.</i> , (1996)
Ex 45	B2.1	c.12384C>T g.51040C>T	A/V4058	1/96 (1.0%)	Constantinides <i>et al.</i> , (1997)

\* Frequency calculated as affected alleles divided by number of chromosomes tested.

\*\*Genomic sequence lists G at nucleotide 24474, however cDNA has A at the corresponding position 2427

\*\*\* B78.1 was homozygous for I/V4044 producing a different SSCP pattern compared to heterozygotes for this change.

In this study, the non-pathogenic missense variations were investigated in greater detail in terms of their predicted effect on protein topology. The analysis was conducted using nnPredict, a secondary protein prediction tool (<http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html>). To assess evolutionary conservation, the amino acid involved was compared between human and mouse and, when possible, the Japanese Pufferfish (<http://www.sanger.ac.uk/Software/Pfam/index.html>). The type of amino acid change was also determined. The results are presented below.

### **Q/R 739**

Glutamine, a polar amino acid with an amide side chain, is replaced by arginine, a basic amino acid. The glutamine residue is conserved between: human (Accession # L33242) and mouse (O08852), however another published human polycystin-1 sequence (Accession # P98161) lists this residue as arginine. One can presume therefore that the variation between published sequences represents a non-disease causing polymorphism. The change occurs between the LDL-A domain and the second PKD repeat. The Q/R 739 mutation causes a marked change in the predicted secondary protein structure, however it is in a region of the protein which does not correspond to any known functional domain. Q/R 739 was detected in one Australian patient (A42.1) and subsequently in an ADPKD family from Bulgaria (personal communication, Dr. Nadja Bogdanova).

**Q739R**

Human (L33243) AGPGALLHCSPAPGHPGP**Q**APYLSANASSW  
Human (P98161) AGPGALLHCSPAPGHPGP**R**APYLSANASSW  
Mouse (O08852) LLQCPLASSC PG **Q**ALYLSTNASDW

**Q739**

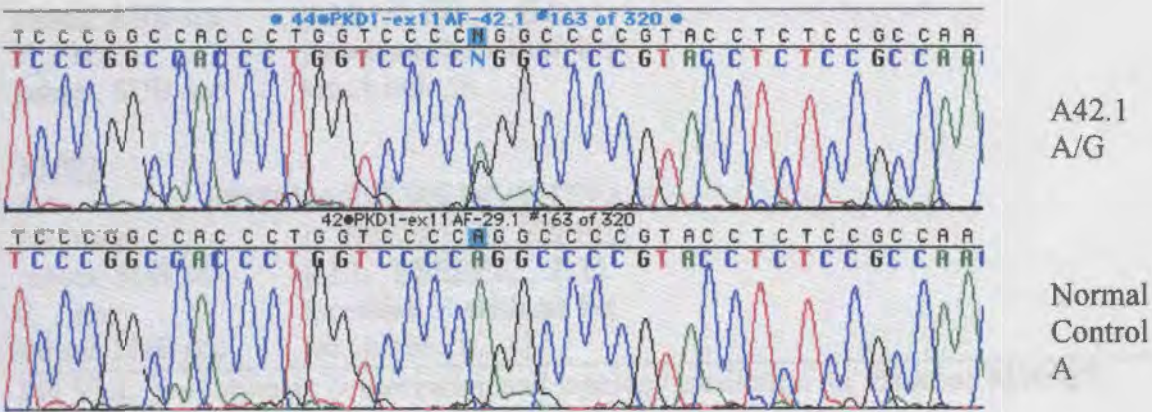
AA |AGPGALLHCSPAPGHPGP**Q**APYLSANASSW|  
PHD sec | EEEE |  
subset: SUB sec |LLL.EEEE.LLLLLLLLLLLL...LLLLLL|  
P\_3 acc |eeeebbbbbbbeee eeebe bbbbeeee|  
subset: SUB acc |e...bbb.b....e.....|

**R739**

AA |AGPGALLHCSPAPGHPGP**R**APYLSANASSW|  
PHD sec | HHH |  
subset: SUB sec |LLL..LLLLLLLLLLLLL.....LLLLLL|  
P\_3 acc |e beebbee eeeee ebbbeeeeeeee|  
subset: SUB acc |e....b.....e....b....ee....|

**Fig VI.1.** Evolutionary conservation and topology prediction for variation Q739R.

**Sequencing gel**





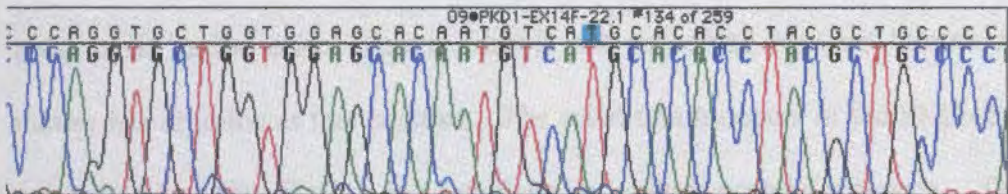
M/T1092

Methionine, a non-polar amino acid with a sulphur-containing side chain, is replaced by threonine, a polar amino acid with an aliphatic hydroxyl side chain. This residue is not conserved between human and mouse, in fact the mouse has threonine at this homologous position. The change occurs in the fourth PKD repeat domain of polycystin-1 and results in predicted changes in secondary protein structure. The novel change was detected in the DNA of three from 17 screened Australian patients (A6.1, A29.1, A46.1). In two of these families (A6.1 and A46.1), other sequence variants segregating with the disease phenotype have been detected (Chapter V).

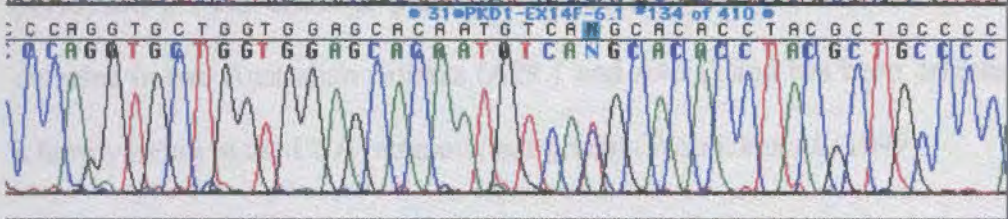
M1092T	
Human	PDPSVAQVLVEHNV <u>M</u> HTYAA
Mouse	PDPTVAQVLVEHNT <u>T</u> HTYTT
M1092	
AA	PDPSVAQVLVEHNV <u>M</u> HTYAA
PHD sec	HH <u>H</u> EEEEEE EEEEE
subset: SUB sec	LLL..H..EE.....E.L
P_3 acc	eeeebbebbbbebbbebb be
subset: SUB acc	ee...b.bbb...b.....
T1092	
AA	PDPSVAQVLVEHNV <u>T</u> HTYAA
PHD sec	EEEEEEEEE EEEEE
subset: SUB sec	LLL...EEEEEE.....E.L
P_3 acc	eeeebbebbbbebbbebb be
subset: SUB acc	ee...b.bbb...b.....

Fig VI.2. Evolutionary conservation and topology prediction for variation M1092T.

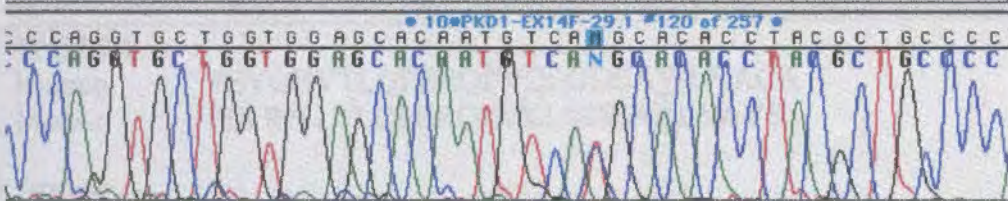
Sequencing Gel M/T1092



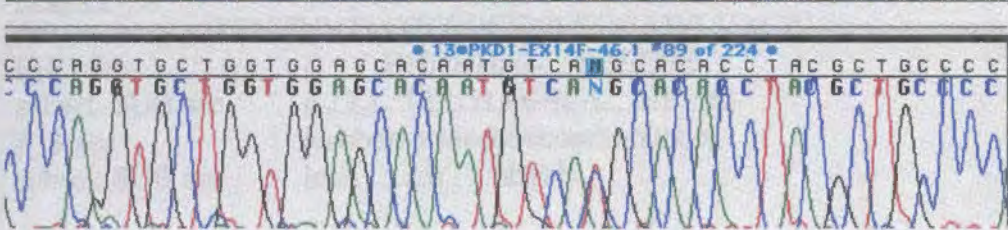
Normal  
Control



A6.1



A29.1



A46.1

W/R 1399

Tryptophan, a non-polar amino acid with an aromatic side chain, is replaced by arginine, a polar basic amino acid. This residue is not conserved between human and mouse, the mouse has arginine at this position. The substitution occurs in the PKD-repeat domain of polycystin-1. No effect is predicted on secondary protein structure. The change was detected in two Australian patients (A29.1 and A46.1) and has been detected previously in a family living in the USA (ethnicity not given) (Watnick *et al.*, 1999).

W1399R

Human	PEVGNVTLQPERQFVQLGDEA <u>W</u> LVACA
Mouse	PEI RN I TLQPERQFVKLGDEA <u>R</u> LVAYS

W1399

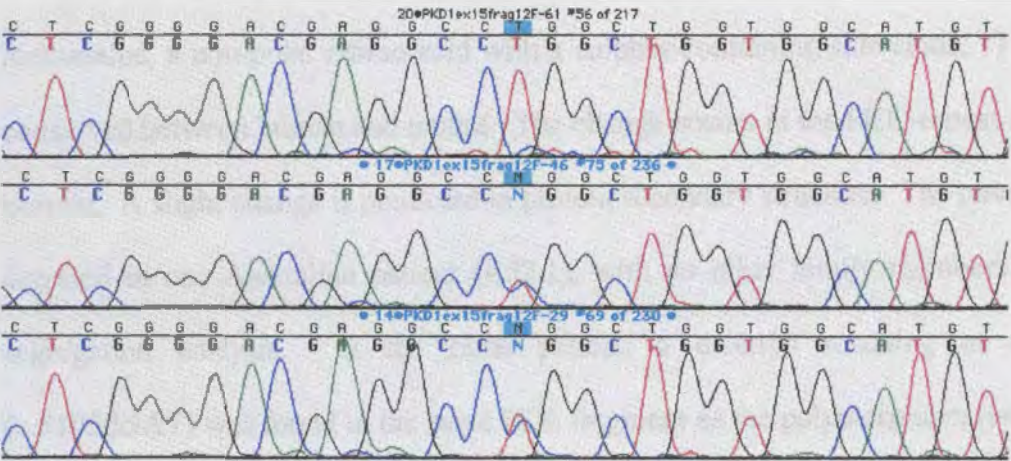
AA	PEVGNVTLQPERQFVQLGDEA <u>W</u> LVACA
PHD sec	HHHHHH EEEEE
subset: SUB sec	LLLL..LLL.H.HHH..L...EEE.L
P_3 acc	eebb <b>b</b> b eee <b>b</b> bebeee <b>b</b> bbbbb
subset: SUB acc	ee.....b.....b <b>b</b> bb..

R1399

AA	PEVGNVTLQPERQFVQLGDEA <u>R</u> LVACA
PHD sec	HHHHHE EEEEE
subset: SUB sec	L.LLLLLLL...H...L...EEEEL
P_3 acc	eebb <b>b</b> b eee <b>b</b> bebeee <b>b</b> bbbbb
subset: SUB acc	ee.....b.....b <b>b</b> bb..

Fig VI.3. Evolutionary conservation and topology prediction for variation W1399R.

Sequencing Gel W/R1399



Normal  
Control

A46.1

A29.1

This deletion was investigated as being due to a deletion of the PCR fragments with the TA-PCR method (1998). Subsequent sequence analysis of the cloned fragments demonstrated that the 100 bp deletion was due to a deletion of the 100 bp fragment.

Protein  
Human  
Mouse

Protein  
Human  
Mouse

Protein  
AA  
Protein  
Human  
Mouse

Protein  
Human  
Mouse

Protein  
AA  
Protein  
Human  
Mouse

Protein  
Human  
Mouse

Fig. 1. Comparison of the sequence and the size of the protein for the human T104H1.

**T/M 1649**

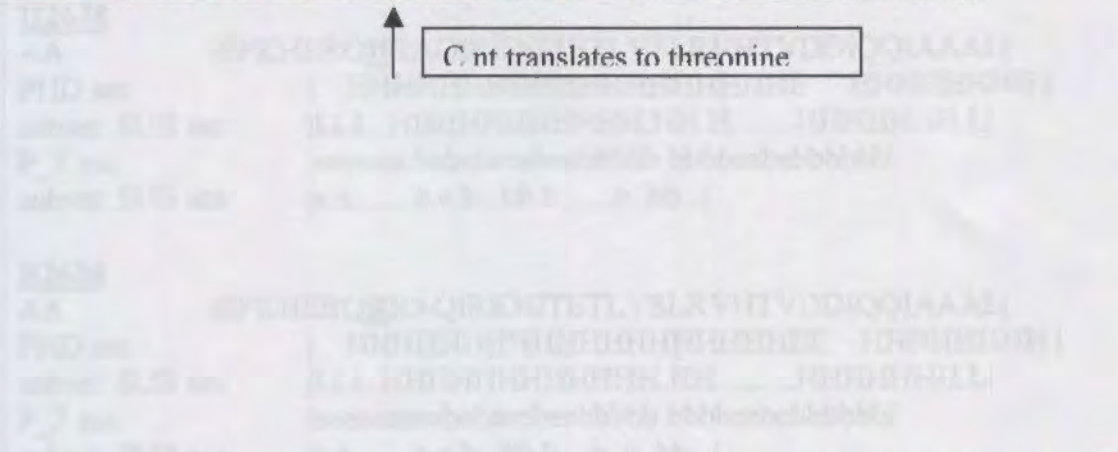
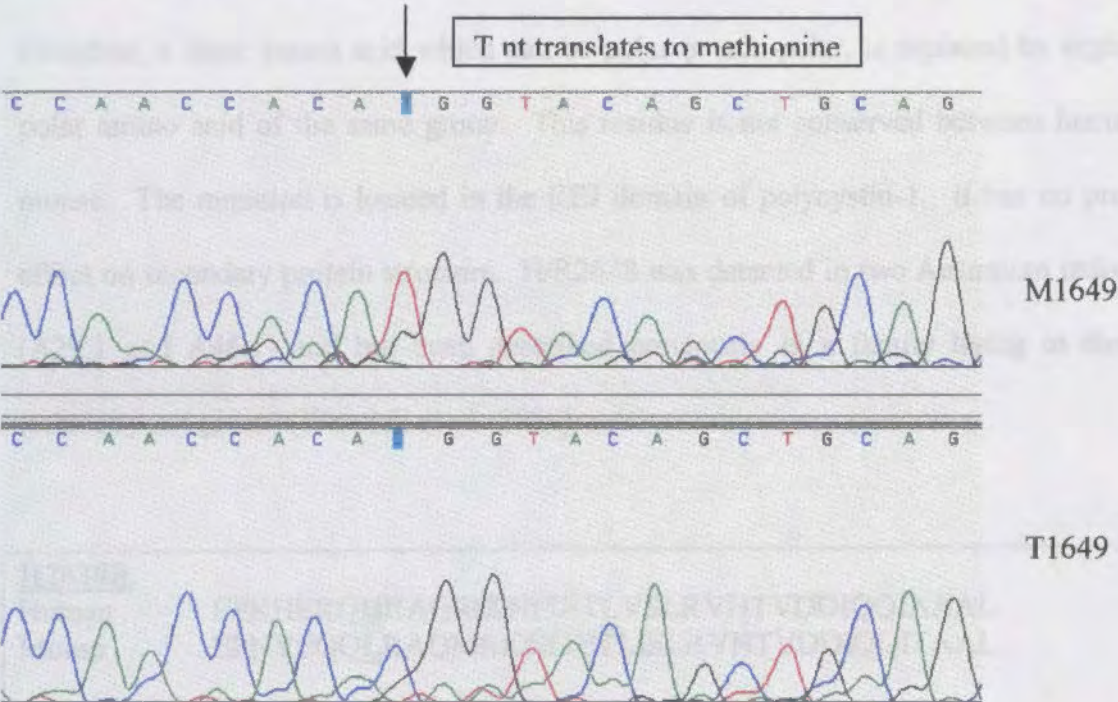
Threonine, a polar amino acid with an aliphatic hydroxyl side chain, is replaced by methionine, a non-polar amino acid with a sulphur-containing side chain. This residue is conserved between human and mouse. The change occurs in the PKD-repeat domain of the protein. A slight change is predicted in protein secondary structure. The novel change was detected in one Australian patient (A52.1), with no other family members available for segregation analysis. In the same patient, a deletion resulting in a frame-shift (c.5105delAT) was found in the same PCR fragment as the polymorphism (see Chapter V). This deletion was investigated in detail after cloning of the PCR fragments with the TA-  
 . 'O cloning kit (Invitrogen). Subsequent sequence analysis of the cloned fragments demonstrated that the T/M 1649 variation and the deletion occurred on different chromosomes.

<b>T1649M</b>	
Human	PTNH <b>T</b> VQLQAVVRDGTNVSYSWTAWR
Mouse	PTNY <b>T</b> LQLQAAVRDGTN I SYSWTAQQ
<b>T1649</b>	
AA	PTNH <b>T</b> VQLQAVVRDGTNVSYSWTAWR
PHD sec	EEEEEEEEE EEEEEEEEE
subset: SUB sec	LLLL.EEEEE..LLL.E.EE..EEL
P 3 acc	eeebbebbbbeebeebbebbbbe
subset: SUB acc	e....b.b..b.....b..b.....
<b>M1649</b>	
AA	PTNH <b>M</b> VQLQAVVRDGTNVSYSWTAWR
PHD sec	EEEEEEEEE EEEEEEEEE
subset: SUB sec	LL....E.EEEE.LLL...EEE...L
P 3 acc	eee ebbbebbb ebeebbbbbe
subset: SUB acc	ee....b.b.bb.....

**Fig VI.4.** Evolutionary conservation and topology prediction for variation T1649M.



Electrophoretogram T/M1649



**H/R2638**

Histidine, a basic amino acid which can be polar or non-polar, is replaced by arginine, a polar amino acid of the same group. This residue is not conserved between human and mouse. The mutation is located in the REJ domain of polycystin-1. It has no predicted effect on secondary protein structure. H/R2638 was detected in two Australian individuals (A29.1 and A46.1) and has been described previously in a family living in the USA (ethnicity not given) (Watnick *et al.*, 1999).

**H2638R**

Human	EPKHERQ <b>H</b> RAQIRKNITETLVSLRVHTVDDIQQIAAAL
Mouse	EPNVEQ <b>Q</b> LRAQMRKNITETLISLRVNTVDDIQQITAAL

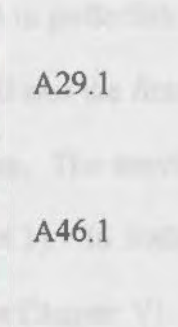
**H2638**

AA	EPKHERQ <b>H</b> RAQIRKNITETLVSLRVHTVDDIQQIAAAL
PHD sec	HHHHHHHHHHHHHHHHHHHHE HHHHHHHH
subset: SUB sec	LLL..HHHHHHHHHHHH.HH.H.....HHHHHHH.L
P_3 acc	eeeeeebebeeebeebbbb bbbbeebbbbbbb
subset: SUB acc	e.e.....b.e.b...bb.b.....b..bb...

**R2638**

AA	EPKHERQ <b>R</b> RAQIRKNITETLVSLRVHTVDDIQQIAAAL
PHD sec	HHHHHHHHHHHHHHHHHHHHEE HHHHHHHH
subset: SUB sec	LLL.HHHHHHHHHHHHH.HH.....HHHHHHH.L
P_3 acc	eeeeeebebeeebeebbbb bbbbeebbbbbbb
subset: SUB acc	e.e.....b.e.b...bb.b....b..b..bb...

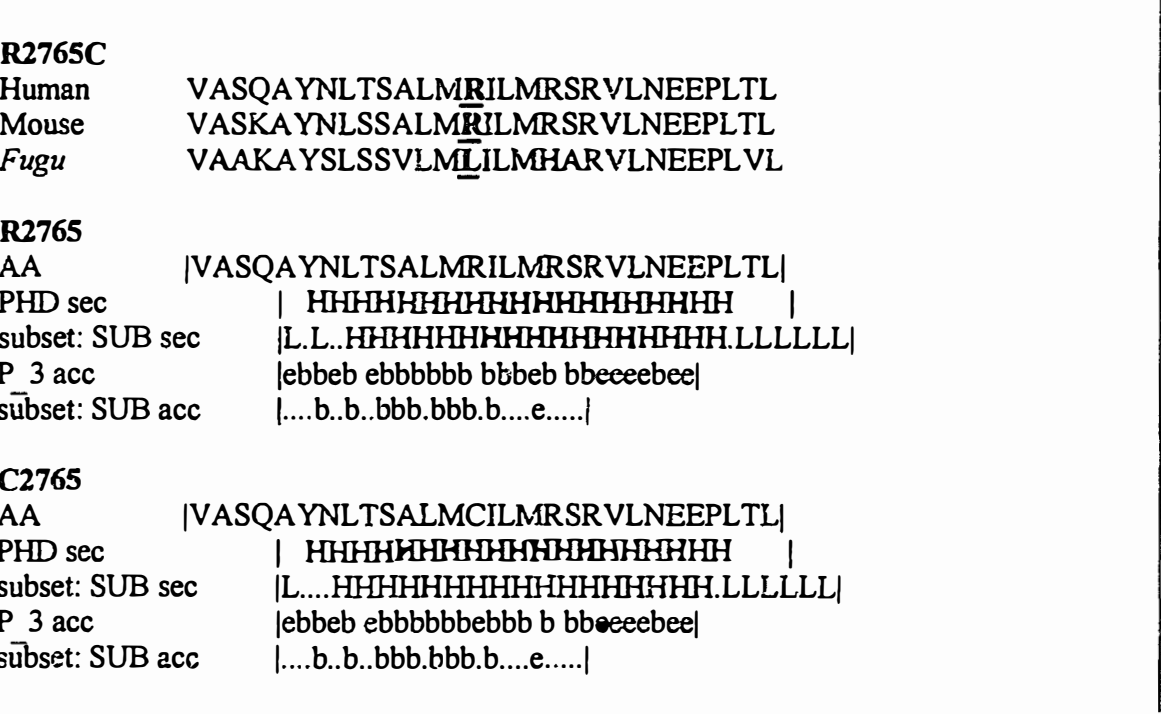
**Fig VI.5.** Evolutionary conservation and topology prediction for variation H2638R.





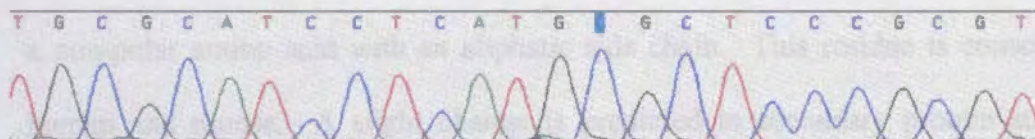
**R/C2765**

Arginine, a polar basic amino acid, is replaced by cysteine, a polar amino acid with a sulphhydryl side chain. This residue is conserved in human and mouse but not in pufferfish. The change occurs in the extracellular region of the protein, between the REJ and the first transmembrane domain. No effect is predicted on secondary protein structure. The novel R/C 2765 variant was detected in two Australian patients (A42.1 and A46.1). In both families, probable disease-causing mutations in *PKD1* have been detected (see Chapter V).

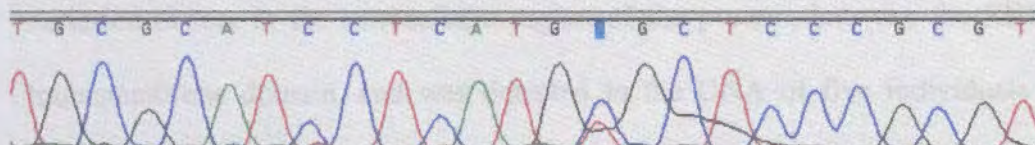


**Fig VI.6.** Evolutionary conservation and topology prediction for variation R2761C.

# Sequencing Gel R/C2765



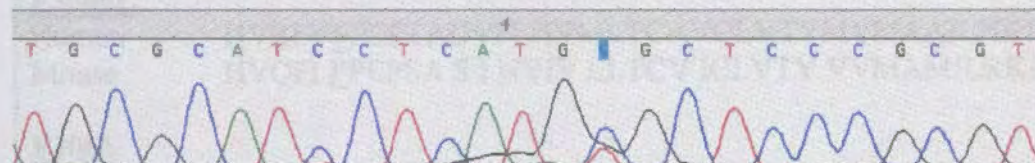
Normal  
Control



A29.1



Normal  
Control



A46.1

**F/L3066**

Phenylalanine, a non-polar amino acid with an aromatic side chain, is replaced by leucine, a non-polar amino acid with an aliphatic side chain. This residue is conserved between human and mouse. A slight change is predicted in secondary protein structure. The variation occurs in the extracellular region of the protein, between the REJ and the first transmembrane domain, and was detected in the DNA of five individuals (A6.1, A9.1, A29.1, A36.1 and A46.1). It had been detected previously in a British population at a frequency of 23.5% (Peral *et al.*, 1997).

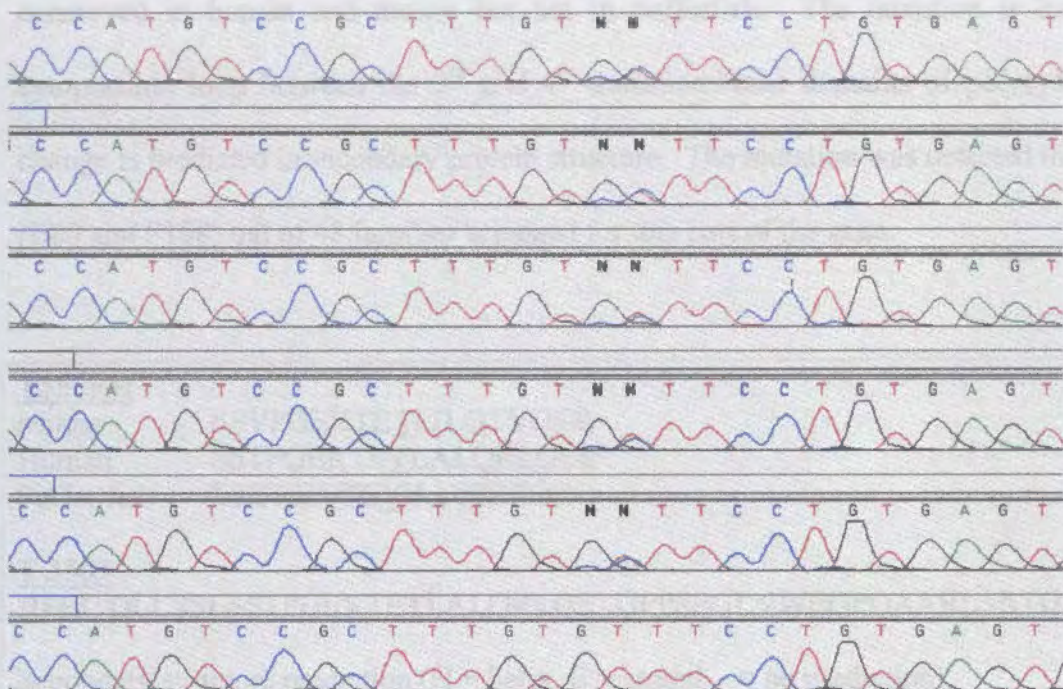
**F3066L**  
Human        HVRFVFPEPTADVNYIVMLTCAVCLVTYMVMAAILHKLDQLD  
Mouse        HVQFI FPEPSA S I NYIV LLTCV ICLVTY VVMAMILRKLDQLD

**F3066**  
AA            |HVRFVFFPEPTADVNYIVMLTCAVCLVTYMVMAAILHKLDQLD|  
PHD sec        | EEEEE    EEEEE    EEEHHHHHHHHH |  
subset: SUB sec | LEEEE.LLLLLL..EEEEEE.LLL.EE...HHHHHHHH.LLL|  
P\_3 acc        | eb bbbeebeeebebbbbbbeeebbbbbbeeebeeebe|  
subset: SUB acc | ...bb.....b.bbbb....b..bbbbbb.b....|

**L3066**  
AA            |HVRFVLLPEPTADVNYIVMLTCAVCLVTYMVMAAILHKLDQLD|  
PHD sec        | EEEE    EEEEE    EEEHHHHHHHHHHH |  
subset: SUB sec | L.EEE.LLLLLL..EEEEEE.LL..E...HHHHHHHH.LLL|  
P\_3 acc        | eb bbbeebeeebebbbbbbeeebbbbbbeeebeeebe|  
subset: SUB acc | .b.bbb.....e...bbbb.bb....b..bbbbbb.....|

**Fig VI.7.** Evolutionary conservation and topology prediction for variation F3066L.

# Sequencing Gel F/L3066



A6.1

A9.1

A29.1

A36.1

A46.1

Normal  
control

**T/M 3509**

Threonine, a polar amino acid with an aliphatic hydroxyl side chain, is replaced by methionine, a non-polar amino acid with a sulphur-containing side chain. This residue is conserved in human and mouse but not in pufferfish. The mutation is located in a cytoplasmic loop between the 3<sup>rd</sup> and 4<sup>th</sup> transmembrane domains of polycystin-1. No change is predicted in secondary protein structure. The mutation was detected in 2 families (B80 and P108) out of 48 families screened for this part of the gene.

**T3509M**

Mouse	SSVPGEKTET <u>L</u> LILQTVGEE
Human	SSTPGEKTET <u>L</u> LALQRLGEL
Puffer fish	SSIFEDQTEV <u>L</u> LLQKHHEP

**T 3509**

HMETDLLSSLSTPGEKTETLLALQRLGELGPPSPGLNWEQPQAARLSRTG

Secondary structure prediction (H = helix, E = strand, - = no prediction):  
----HHH-----HHHHHHH-----HHHHH----

**M 3509**

HMETDLLSSLSTPGEKTEMLALQRLGELGPPSPGLNWEQPQAARLSRTG

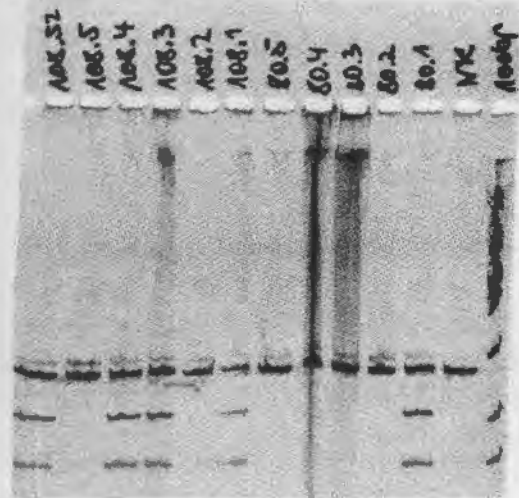
Secondary structure prediction (H = helix, E = strand, - = no prediction):  
----HHH-----HHHHHHHH-----HHHHH----

**Fig VI.3.** Evolutionary conservation and topology prediction for variation T3509M.

**Exon 35**

**SfNaI Restriction digest**

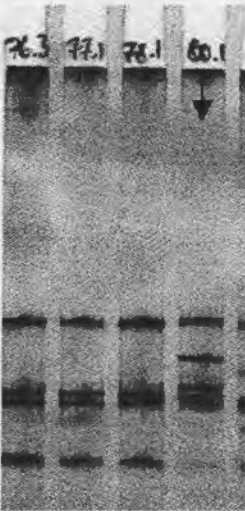
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13



Lanes 1-11 Families P108 & B80  
Lane 12 Normal Control (NC)  
Lane 13 Size Marker

**SSCP gels arrows indicate B80.1 & P108**

1 2 3 4



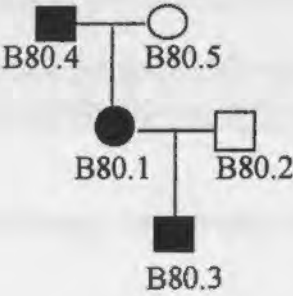
Lanes 1-3 NC  
Lane 4 B80.1

1 2 3



Lane 1 P108.1  
Lane 2 NC  
Lane 3 Size Marker

**Pedigree B80**





**S/I 3592**

Serine, a polar amino acid with an aliphatic hydroxyl side chain, is replaced by isoleucine, a non-polar amino acid with an aliphatic side chain. This residue is conserved in human and mouse but not in pufferfish. The mutation occurs in the 4<sup>th</sup> transmembrane domain of the protein. A slight change is predicted in secondary protein structure. Two families (B1, A29), from 48 were detected to have this change.

**S3592I**

Mouse	WLLSSSSSFLASFLGWEP
Human	WLLSS <u>S</u> ASFLASFLGWEP
Pufferfish	WLLSC <u>F</u> ASFLCSCLLLEP

**S 3592**

PPGVSVAWLLSSSASFLASFLGWEPLKVLLEALYFSLVAKRLHPDEDDTL

Secondary structure prediction (H = helix, E = strand, - = no prediction):

---EEEEEH--HH~~EEEE~~-----HHHHHHHHHHHHHHHHHHHH-----

**I 3592**

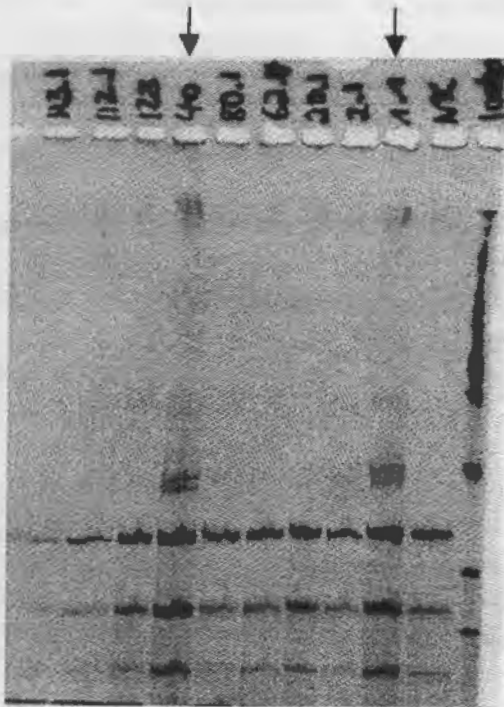
PPGVSVAWLLSSIASFLASFLGWEPLKVLLEALYFSLVAKRLHPDEDDTL

Secondary structure prediction (H = helix, E = strand, - = no prediction):

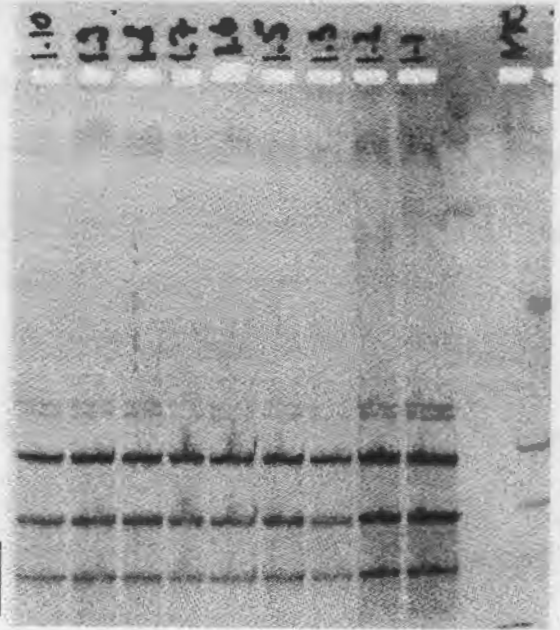
---EEEEHHHHHHHHHHHH-----HHHHHHHHHHHHHHHHHHHH-----

**Fig VI.9.** Evolutionary conservation and topology prediction for variation S3592I.

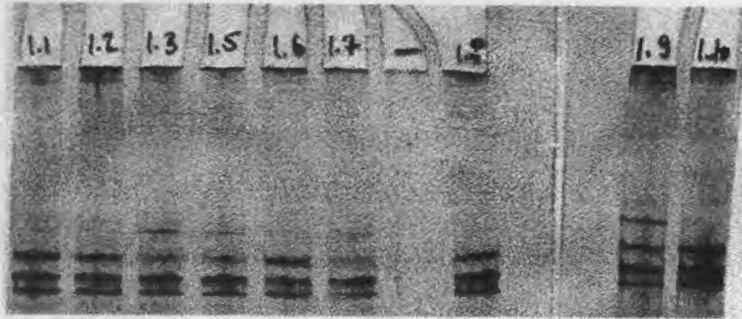
**Restriction Digests of Exon 36 (HhaI);** Gel A, arrows indicate A29.1 and B1.1  
Gel B Family B1 with normal control (far right)



A

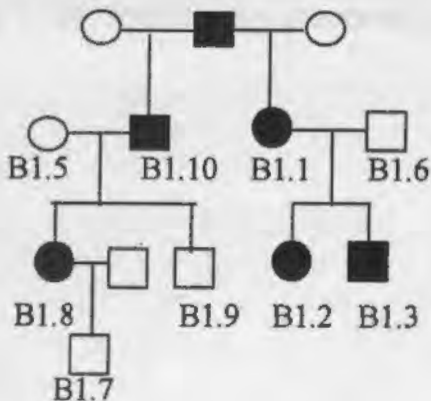


B



SSCP Family B1  
Exon 36  
B1.1 – B1.10 ran in  
lanes left to right.  
Aberrant SSCP does  
not co-segregate with  
disease as seen in  
pedigree below

**Pedigree B1**





G/Q 3695

Glycine, a polar amino acid, with an aliphatic side chain consisting of a hydrogen atom, is replaced by glutamine, a polar amino acid with an amide group side chain. This residue is conserved between human and mouse but not in pufferfish. The variation occurs in an extracellular loop of the protein, between the 6<sup>th</sup> and 7<sup>th</sup> transmembrane domains. A slight change is predicted in secondary protein structure. One family (P109) out of 48 was found to have this change.

**G3695Q**

Mouse	DASCH <b>G</b> HAYRLQSAIKQEL
Human	DASCH <b>G</b> HAYRLQSAIKQEL
Puffer fish	DSDKD <b>A</b> HSLRLRTQLQRAL

**G 3695**

LLVTLLASYGDASCH**G**HAYRLQSAIKQELHSRAFLAITRS

Secondary structure prediction (H = helix, E = strand, - = no prediction):

--EEHH-----HHHHHHHHHHHHHHHHHHHH--

**H 3695**

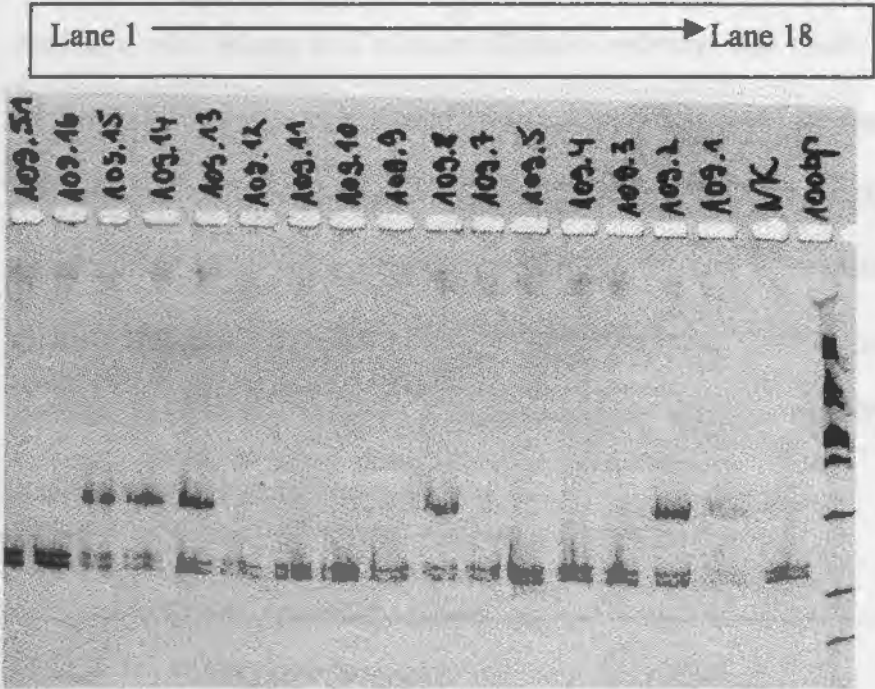
LLVTLLASYGDASCHQ**H**AYRLQSAIKQELHSRAFLAITRS

Secondary structure prediction (H = helix, E = strand, - = no prediction):

--EEHH-----HHHHHHHHHHHHHHHHHHHHHHHHHHHH--

**Fig VI.10.** Evolutionary conservation and topology prediction for variation G3695Q.

Restriction Digest NcoI of Exon 38



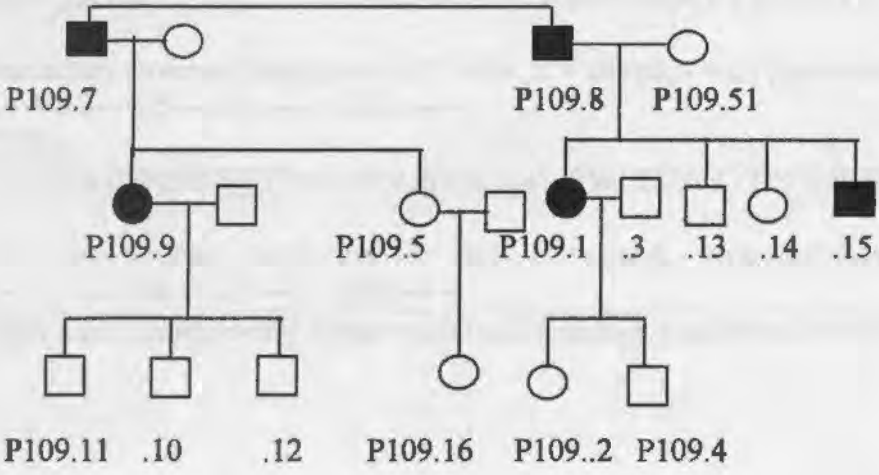
Restriction Digest  
Family segregation  
P101

Lanes 1-16  
Family P101

Lane 17  
Normal Control

Lane 18  
1000bp Marker

Pedigree



**S/T 3791**

Serine, a polar amino acid with an aliphatic hydroxyl side chain, is replaced by threonine, an amino acid of the same group. This residue is conserved between human and mouse and no information is available on pufferfish. The sequence variation corresponds to an extracellular loop in the protein between the 6<sup>th</sup> and 7<sup>th</sup> transmembrane domains. No change is predicted in secondary protein structure. One family (P101) from 48 was detected to have this change. A disease-causing mutation (S3770X) had been detected in this family in the same amplicon.

**S3791T**

Mouse	WQSVVQNG <u>S</u> ET
Human	WESP---HNG <u>S</u> GT

**S 3791**

YDVGWESPHNGSGTWAYSAPDLLGAWSWGSCAVYDSGGYV

Secondary structure prediction (H = helix, E = strand, - = no prediction):

-----EE----H-----EEE-----

**T3791**

YDVGWESPHNGTGTWAYSAPDLLGAWSWGSCAVYDSGGYV

Secondary structure prediction (H = helix, E = strand, - = no prediction):

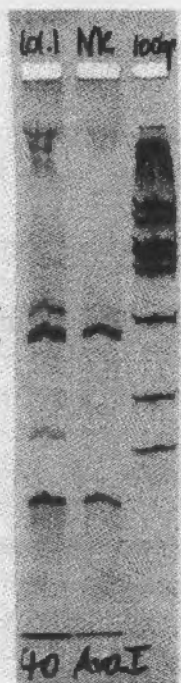
-----EE----H-----EEE-----

**Fig VI.11.** Evolutionary conservation and topology prediction for variation S3791T.

1 2 3

Ava I restriction digest Exon 40

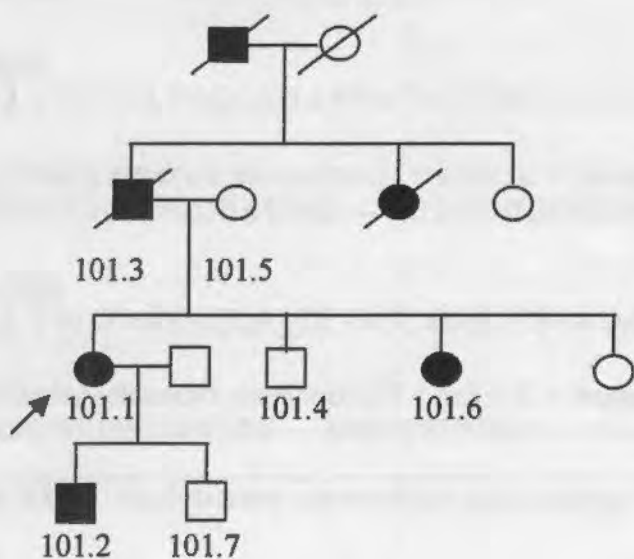
Lane 1 P101.1  
Lane 2 Normal Control  
Lane 3 Size Standard (100bp ladder)



264bp  
182bp

82bp

Pedigree:



**I/V 4044**

Isoleucine, a non-polar amino acid with an aliphatic side chain, is replaced by valine, an amino acid of the same group. This residue is conserved between human and mouse and no information is available on pufferfish. The variation is located in the 9<sup>th</sup> transmembrane region of the protein. No change is predicted in secondary protein structure. Of 48 families screened, 21 were detected with this polymorphism. Family analysis was not undertaken for this common change. I/V 4044 has been described previously (Rossetti *et al.*, 1996) as a non-pathogenic polymorphism in an Italian family. The I4044 variant was detected at a frequency of 26% in the Spanish population (Badenas *et al.*, 1999).

**I4044V**

Mouse	LGVA YAQMA <u>I</u> LLISSGAD
Human	LGVA YAQLA <u>I</u> LLVSSCVD

**I 4044**

LGLVVLGVAYAQLAILLVSSCVDSLWSVAQALLVLCPGTGLSTLCPAESW

Secondary structure prediction (H = helix, E = strand, - = no prediction):

--EEEEHHHHHHHHHHEE-----HHHHHHHHHHEE-----E-----

**V 4044**

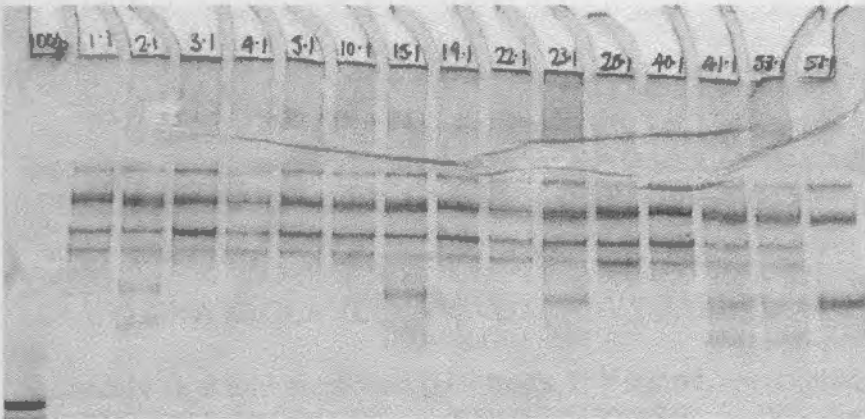
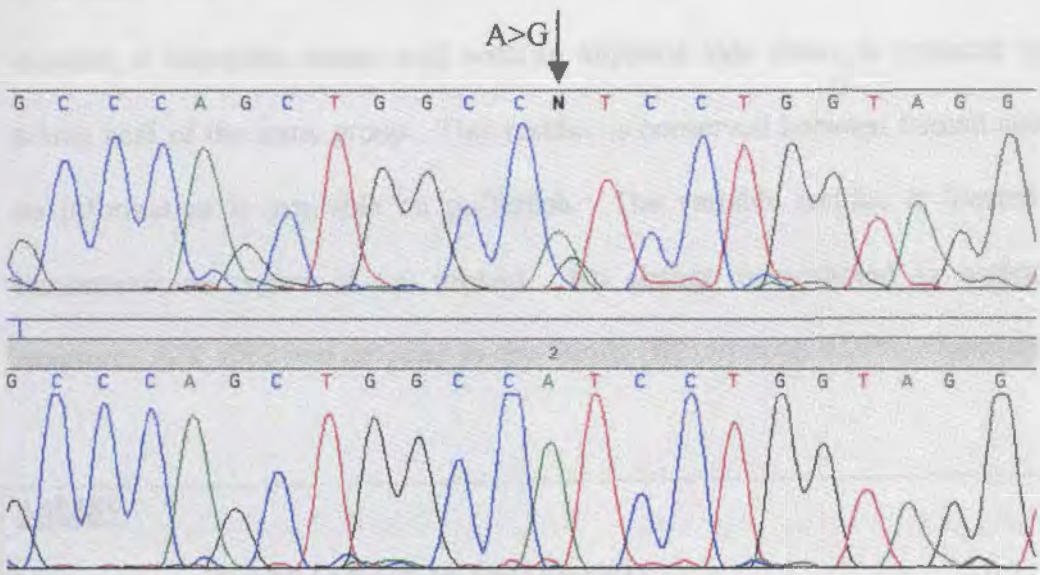
LGLVVLGVAYAQLAVLLVSSCVDSLWSVAQALLVLCPGTGLSTLCPAESW

Secondary structure prediction (H = helix, E = strand, - = no prediction):

--EEEEHHHHHHHHHHHEE-----HHHHHHHHHHEE-----E-----

**Fig VI.12.** Evolutionary conservation and topology prediction for variation I4044V.

Sequencing Gel I/V4044



SSCP Exon 44

This common polymorphism can be seen across the gel with its two distinct SSCP patterns. Lane marked 57.1 (far right) represents a sample that was homozygous for the less common allele.

**A/V 4058**

Alanine, a non-polar amino acid with an aliphatic side chain, is replaced by valine, an amino acid of the same group. This residue is conserved between human and mouse and no information is available on pufferfish. The variable residue is located in the 10<sup>th</sup> transmembrane region of the protein. No change is predicted in secondary protein structure. A/V 4058 was detected in one family (B2) from 48 ADPKD families studied.

**A4058V**

Mouse	TLYNM <u>A</u> RAFLVLCPGARVPTL
Human	SLWSV <u>A</u> QALLVLCPGTGLSTL

**A 4058**

LGLVVLGVAYAQLAVLLVSSCVDSLWSVAQALLVLCPGTGLSTLCPA<sup>E</sup>SW

Secondary structure prediction (H = helix, E = strand, - = no prediction):  
--EEEEHHHHHHHHHHHHEE-----HHHHHHHHHLEE-----E-----

**V 4058**

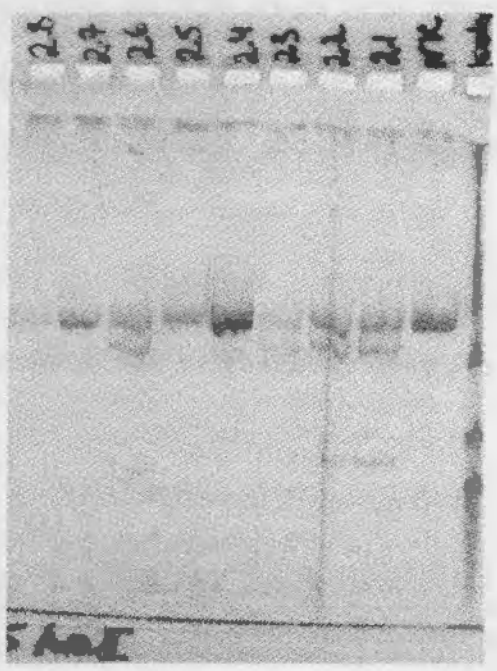
LGLVVLGVAYAQLAVLLVSSCVDSLWSVVQALLVLCPGTGLSTLCPAESW

Secondary structure prediction (H = helix, E = strand, - = no prediction):  
--EEEEHHHHHHHHHHHHEE-----EHHHHHHHEEEE-----E-----

**Fig VL13.** Evolutionary conservation and topology prediction for variation A4058V.

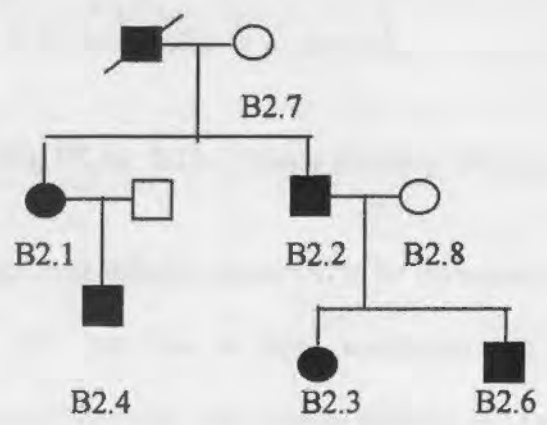
# **Restriction Digest for family B2 (AvaII) Exon 45**

Lanes 1-8: Family B2  
 Lane 9: Normal Control  
 Lane 10: Size standard (cut off picture)



**SSCP Exon 45**  
 1. Normal control  
 2. B2.1

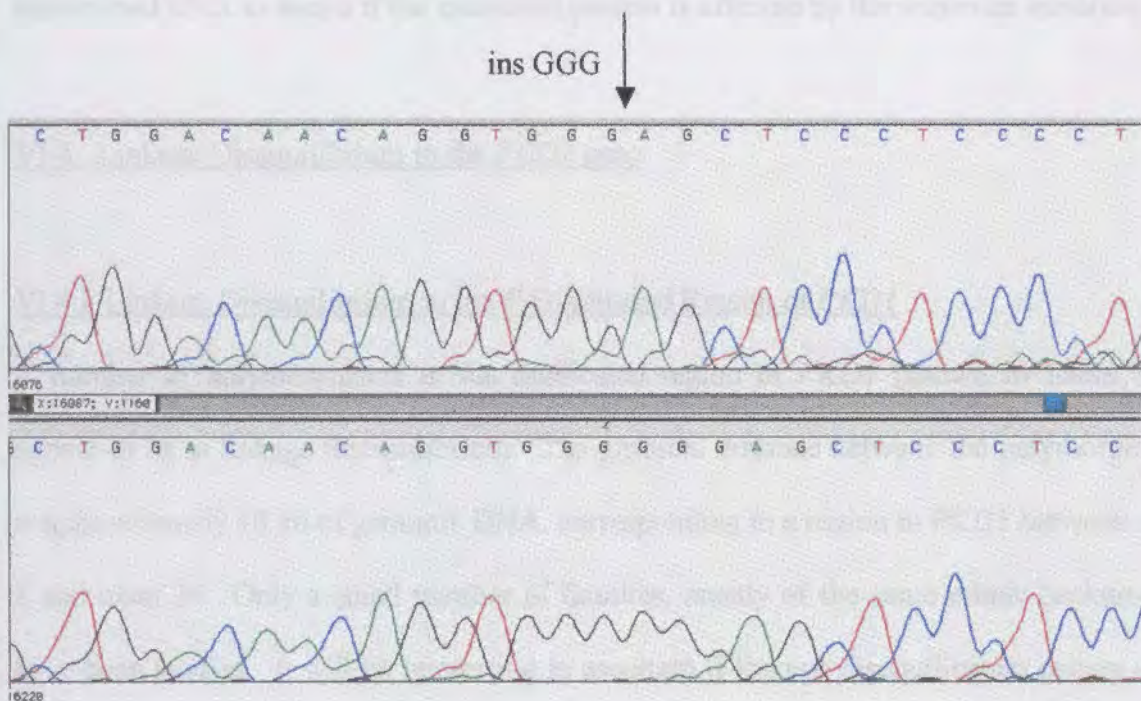
## **Pedigree B2**





### VI.3. A variant in the 3' Unique Region segregating with the disease phenotype

An intronic insertion was found to segregate with the ADPKD phenotype in Bulgarian family B23, and was characterised using the TOPO-TA cloning kit (Invitrogen). Clones containing the PCR inserts from patient B23.1 were subject to sequence analysis. The resulting electrophoretogram below (Fig VI.14) displays the sequence variation between the cloned PCR insert of the wild-type allele and the IVS41+5 ins3 variant allele.



**Fig VI.14.** B23.1 Clones showing Wild-type allele and below the IVS41 +5insggg

The change was initially suspected to be pathogenic as it segregated with the disease in this family. The insertion of three nucleotides so close to the donor splice site could hypothetically interfere with normal splicing and thus represent a disease-causing mutation. This sequence change had been suggested previously to be the disease-causing mutation in a French family (Perrichot *et al.*, 2000a). However, shortly after the publication of the

French report, Koptides *et al.* (2000) showed that the same DNA variation was detected in a Greek-Cypriot kindred, where it did not segregate with the disease. In this study, the mutation occurred in all four affected subjects in family B23 and was not found in one unaffected family member. No disease-causing mutation has been identified in this family. The change was not detected in the other 47 families screened. RNA from this family was not available for analysis. It was therefore not possible to determine the sequence of the transcribed RNA to assess if the translated protein is affected by the sequence variation.

#### VI.4. Linkage Disequilibrium in the *PKD1* gene

##### VI.4.i. Linkage Disequilibrium in the 5' Duplicated Region of *PKD1*

A number of polymorphisms in the duplicated region of *PKD1* (shown in Table VI.3) appear to be in linkage disequilibrium. The physical distance between the polymorphisms is approximately 18 kb of genomic DNA, corresponding to a region in *PKD1* between exon 5 and exon 26. Only a small number of families, mostly of the same ethnic background, have been studied. It will be interesting to ascertain if linkage disequilibrium occurs in all ethnic groups and whether this SNP haplotype can be used for investigating affinities between populations.

**Table VI.3.** Shared polymorphisms of the 5' Duplicated region of *PKDI*

Patient	A6	A29	A46	A9	A13	A36	A7	A10
Seq. Change								
L373	na			+	+	+		
IVS9del7				+	+	+		
P900				+	+	+		
P910				+	+	+		
G1021	+	+	+					
L1037							+	+
M/T1092	+	+	+					
A1124	na		+	+				
S1125	na		+	+				
W/R1399	na	+	+					
A1555	na	+	+	+	+	+		
T1558	na						+	+
T1724	na	+	+	+	+	+		
L1921	na			+	+	+		
L2389	na		+	+	+	+		
L2481	na	+	+		+	+		
H/R2638	na	+	+					
F/L3066	+	+	+	+		+		
P3110	+	+	+	+		+		

\*na Patient A6 was not screened in the region from exon 1 to 5 and 15 to 21.

#### VI.4.ii. Linkage Disequilibrium in the 3' Unique Region of *PKD1*.

An interesting observation on the inheritance of ancestral chromosomes was related to the co-existence of three common polymorphisms in the final three exons of *PKD1*, over a physical distance of 670 bp. These are I/V4044 in exon 44; c.12484A>G (A4091) in exon 45 and c.12838T>C (in exon 46). Of 85 *PKD1*-linked individuals from three ethnic groups screened in this study, 29 (34%) shared all three polymorphisms and one (B78.1) was homozygous for each. Patients who had the polymorphisms included 19 of 57 Australians (33%), 2 of 7 Poles (29%), and 8 of 21 Bulgarians (29%), suggesting that this haplotype is common across Europe.

Greater variation in linkage disequilibrium was observed in Bulgarian individuals, where five had a combination of one or two of the polymorphisms but not all three (see Table VI.4). Two additional polymorphisms in exon 45, namely (A/V4058 and c.12617C>T 4220A), were noted to occur in addition to the other three polymorphisms. The sub-group of individuals with all five polymorphisms represents nine of the total of 55 patients (16%). This linkage disequilibrium had previously been noted in a Spanish population (Badenas *et al.*, 1999): 26% of the Spanish population had all three common polymorphisms, and 8% of individuals also had the additional A/V4058 polymorphism. The ease of investigating the I/V4044/c.12484A>G/c.12838T>C haplotype located in the unique region of *PKD1*, its high frequency in diverse ethnic groups, and the observed differences, suggest that these polymorphisms may be useful in population genetics research.

**Table VI.4.** The shared common polymorphisms of *PKD1* exons 44-46

<b>Individuals with Exon 44 I/V4044</b>	<b>Individuals with Exon 45 c.12484A&gt;G 4091A</b>	<b>Individuals with Exon 46 c.12838T&gt;C 4209P</b>
B2.1	B2.1	B2.1
-	-	B3.1
-	B4.1	B4.1
B15.1	B15.1	B15.1
B23.1	B23.1	B23.1
B41.1	B41.1	B41.1
B53.1	B53.1	B53.1
B57.1	B57.1	-
B73.1	B73.1	B73.1
B74.1	B74.1	B74.1
B75.3	B75.3	-
B77.1	B77.1	-
B78.1*	B78.1*	B78.1*
A3.1	A3.1	A3.1
A6.1	A6.1	A6.1
A9.1	A9.1	A9.1
A11.1	A11.1	A11.1
A14.1	A14.1	A14.1
A25.1	A25.1	A25.1
A27.1	A27.1	A27.1
A29.1	A29.1	A29.1
A34.1	A34.1	A34.1
A36.1	A36.1	A36.1
A37.1	A37.1	A37.1
A38.1	A38.1	A38.1
A44.1	A44.1	A44.1
A46.1	A46.1	A46.1
A49.2	A49.2	A49.2
A55.1	A55.1	A55.1
A57.1	A57.1	A57.1
A59.1	A59.1	A59.1
A72.1	A72.1	A72.1
P109.1	P109.1	P109.1
P111.1	P111.1	P111.1

\*B78.1 was homozygous for each polymorphism

## VII. Discussion



## VII. Discussion

During the eight years since the cloning of the *PKD1* gene (The European PKD Consortium, 1994) there has been limited progress in understanding the mutational basis of ADPKD. The detection of mutations in *PKD1* has been hindered by the existence of homologous genes interfering with *PKD1*-specific PCR amplification, as well as by the nature of the *PKD1* gene. Its very high GC content, 62% overall (The European PKD Consortium, 1994), poses significant technical problems with requirements for high annealing temperatures and careful optimisation of PCR amplification conditions. By far the most significant obstacle, however, has been the co-amplification of *PKD1* homologous genes (HGs), leading to a heavy bias of mutation detection studies toward the unique 3' region of *PKD1*.

During the cloning of *PKD1*, three HG transcripts were identified (International PKD Consortium, 1995) whose sequences were not characterised. Since the sequence information on this part of the genome is still incomplete, the structure of these genes and even their exact number have remained unknown. Loftus *et al.* (1999) characterised a BAC clone (GenBank AC002039) containing two *PKD1* homologues, referred to as HG1 and HG2. This publication opened the way to mutation analysis of the entire *PKD1* gene, with subsequent studies using the sequence of BAC clone AC002039 for their PCR primer design (Watnick *et al.*, 1999; Thomas *et al.*, 1999; Rossetti *et al.*, 2001; Phakdeekitcharoen *et al.*, 2000 & 2001). The high degree of homology (about 95%) between *PKD1* and the HGs, has added the requirement for long-range PCR amplification of large DNA fragments to the existing problems of *PKD1* analysis. The number of mutation detection studies in ADPKD has remained

small, therefore both the understanding of the molecular defects, and the possibilities for molecular diagnosis in affected families are still limited.

The present study has added important new data on the *PKD1* homologous genes. Using different experimental approaches, we have identified two new homologues (HG 3 and 4), and detailed database searches have led to the detection of two more (HGs 5 and 6). At this stage, the total number of human duplications remains unclear and will probably only be established after completion of the Human Genome Project when all of the presently existing "gaps" are filled. This uncertainty has to be taken into account in the molecular diagnosis and counselling of ADPKD families.

The known number of *PKD1* homologous genes is large enough to raise questions about their role, and the molecular mechanisms leading to their appearance. Our analysis of the model glioblastoma cell line T98G showed that, although HG3 and HG4 transcripts are produced, they are not present in the translationally active polysomal fraction, *i.e.* they are most probably not translated (Bogdanova *et al.*, 2001). However, suppression of ribosome binding to the HGs cannot be discounted under certain cell conditions, and additional work is required into the expression of the *PKD1* HGs to investigate this possibility. Further support is added to the hypothesis that the HGs are transcribed, but not translated, by the abundance of premature termination signals in HG coding sequences. Thus, even if the HGs were translated they would generate short, most probably non-functional, polypeptides. We have therefore proposed that the *PKD1* homologous genes (HGs 1 to 6) are pseudogenes and should be referred to as *PKDIP1* to *PKDIP6* (Bogdanova *et al.*, 2001).



A pseudogene can be defined as "A DNA sequence which shows a high degree of sequence homology to a nonallelic functional gene but which is itself non-functional" (Strachan & Read, 1999, 2<sup>nd</sup> Ed. p553). There are several types of pseudogene, including those with introns and exons that are probably the result of tandem gene duplication events, and pseudogenes with exonic sequence only that originate from the integration of reverse transcribed cDNA into a new chromosomal location. These types of pseudogenes are conventionally non-processed, *ie.* they are non-functional copies of the functional gene. However, an intermediate class of expressed non-processed pseudogenes also exist which undergo transcription (Strachan & Read, 1999). Immediately after a gene duplication event, it is thought that both the original gene and its homologue have the capacity to be expressed and function if the correct regulatory elements are intact. Through time, the HG may undergo deleterious mutations which prevent translation but still allow transcription of the pseudogene (Strachan & Read, 1999).

The *PKD1* pseudogenes are homologous to the genomic DNA of *PKD1* (Loftus *et al.*, 1999, and this study), and therefore probably result from tandem gene duplication events. Given the fact that the mouse *pkdl* gene is not duplicated (Piontek & Germino, 1999), these events have to be recent in mammalian evolution. It would be interesting to discover what sequence changes have made this part of the human genome particularly unstable.

Some pseudogenes, such as the  $\theta$ -globin gene (*HBQ-1*), are known to express a polypeptide product whose role is unknown, and the gene sequence suggests that it is non-functional (Clegg, 1987). One of the recent theories of gene duplication is the

duplication-degeneration-complementation (DDC) model (Force *et al.*, 1999). Gene duplication events can lead to three outcomes. The first is non-functionalisation, where the duplicate becomes a true non-functioning pseudogene; the second outcome is neo-functionalisation, where the duplicate produces a novel protein product; and the third is sub-functionalisation where the duplicate continues to be expressed at lower levels than normal. The DDC model suggests that the sub-functional copy may become essential for gene function and could account for a significant proportion of preserved duplicate genes (Force *et al.*, 1999).

If the *PKDI* HGs are expressed *in vivo* they may interact with *PKDI* or its protein partners to alter the function of polycystin-1. Even if the HGs are not translated, their transcripts may modify *PKDI* expression in various tissues at different stages of development and pathogenesis. These intriguing possibilities remain to be addressed in future studies.

By using sequence information on all six identified *PKDI* pseudogenes, the design of PCR primers for the specific amplification of the *PKDI* gene has been improved. Screening for ADPKD mutations was performed on over 94% of the *PKDI* coding sequence, spanning exons 2 to 21, 23 to 41 and 43 to 46. This represents 12,163 bp of the overall 12909 bp *PKDI* message (Hughes *et al.*, 1995). So far, only two published studies have reported screening a greater proportion of the entire *PKDI* gene: Rossetti *et al.*, (2001) screened the gene between exons 1 to 8, 15 to 17 and 22 to 26 by direct sequence analysis and exons 8 to 46 using PTT. Phakdeekitcharoen *et al.* (2000 & 2001) analysed the complete duplicated coding region (exons 1 to 34) by SSCP analysis

of genomic DNA. Both studies relied on the sequence of HG1 and HG2 (GenBank AC002039) to design *PKD1*-specific primers in the duplicated region of the gene.

A total of 74 sequence variants in the *PKD1* gene were detected during the course of this study. Fifty of these variants are probably non-pathogenic in nature. It is proposed that 24 sequence changes are likely to be disease-causing mutations, of which twenty-one are novel and three have been previously reported. Of the 50 neutral polymorphisms, 31 have been published previously and 19 are described here for the first time.

A two-stage strategy was employed for the detection of mutations in the unique region of *PKD1*, including SSCP screening and the subsequent sequencing of samples with aberrant electrophoretic banding patterns. The efficiency of SSCP as a mutation screening technique was compared with direct sequence analysis, using DNA samples from 15 Australian patients who had been screened previously for exons 44 and 45 in the unique region. As described previously, (Chapter V), exons 44 and 45 were chosen for this comparison because of their different sized PCR amplified fragments and the relatively high number of sequence variations, both neutral and pathogenic, reported in this area of the *PKD1* gene. The failure of SSCP to distinguish between individuals with the A4091 polymorphism in exon 45, and those with A4091 plus the A/V4058 and L4136 polymorphisms, reflects the limitations of the technique, whose sensitivity is lower for PCR fragment lengths over 300-400bp (Hayashi & Yandell, 1993). Therefore the 396 bp exon 45 fragment was probably too long for efficient mutation detection by SSCP. In contrast, no additional variation was detected in the shorter 292bp fragment that included exon 44 when it was re-screened by direct sequencing.

Sequence analysis of exons 2-21 and 23-33 in 17 Australian *PKDI*-linked patients identified 12 possible disease-causing mutations. This represents a mutation detection rate of approximately 71% over 71% of the coding region of the gene. In the unique region of the gene (exons 34-41 & 43-46), nine disease-causing mutations were observed in 48 *PKDI*-linked individuals using SSCP analysis (detection rate 19% over approx. 24% of the coding region). As the area screened did not include the entire gene, and since different numbers of patients were analysed in the unique and the duplicated regions, it was difficult to estimate the overall mutation detection rate. However, as illustrated in Table VII.1, the approximate figures provided here compare favourably to those reported in similar studies.

Table VII.1 Efficiency of *PKD1* screening

Screening method (Template)	Reference	Mutations detected/tested individuals	Region covered (% of cDNA)
<b>SSCP (DNA)</b>	Koptides <i>et al.</i> , (1998)	1 patient only	ex 23-34 (27%)
	Watnick <i>et al.</i> , (1999)	8/35 (23%)	ex 11-21 (42%)
	Koptides <i>et al.</i> , (2000)	1 patient only	ex 23-34 (16%)
	Peral <i>et al.</i> , (1996b)	2/45-6/150 (4%)	ex 34-46 (25%)
	Peral <i>et al.</i> , (1996a)	1 family only	ex 34-46 (25%)
	Neophytou <i>et al.</i> , (1996)	1/14 (7%)	ex 45 (2%)
	Torra <i>et al.</i> , (1997)	1 family only	ex 43-46 (16%)
	Daniells <i>et al.</i> , (1998a & b)	9/147 (6%)	ex 44-45 (3%)
	Badenas <i>et al.</i> , (1999)	7/175 (4%)	ex 43-46 (16%)
	Afzal <i>et al.</i> , (1999)	4/90 (4%)	ex 37-39,44-45 (6%)
	Kim <i>et al.</i> , (2000b)	6/91 (7%)	ex 34-46 (25%)
	Bouba <i>et al.</i> , (2001)	8/53 (15%)	ex 16 - 34 (23%)
	Phakdeekitcharoen (2000)	13/47 (28%)	ex 13-21 (36%)
	Phakdeekitcharoean (2001)	6/41 (14%)	ex 1-12, 22-34 (40%)
	Present study	9/48 (19%)	ex 34-41, 43-46 (24%)
<b>SSCP &amp; DHPLC</b>	Mizoguchi <i>et al.</i> , (2001)	4/176 (2%)	ex 23-46 (41%)
<b>HDA (DNA)</b>	Watnick <i>et al.</i> , (1997)	8/80 (10%)	ex 23-34 (16%)
	Peral <i>et al.</i> , (1995)	2/130 (1%)	ex 43-46 (16%)
	Turco <i>et al.</i> , (1995)	1/20 (5%)	ex 44 (1%)
	Rossetti <i>et al.</i> , (1996)	1/67 (1%)	ex 44-45 (3%)
	Aguari <i>et al.</i> , (2000)	3/40 (7%)	ex 36-46 (23%)
<b>HDA &amp; SSCP</b>	Turco <i>et al.</i> , (1997)	3/60 (5%)	ex 38,40,45 (4%)
<b>DGGE (DNA)</b>	Perrichot <i>et al.</i> , (2000b)	8/146 (5%)	ex 15-21, 23-34 (50%)
	Perrichot <i>et al.</i> , (1999)	8/146 (7%)	ex 34-46 (25%)
	Perrichot <i>et al.</i> , (2000a)	3/146 (2%)	ex 34-46 (25%)
		(total = 13%)	
<b>PTT (<i>in vitro</i> translated)</b>	Roelfsema <i>et al.</i> , (1997)	6/135 (4%)	ex 15 & 23 (30%)
	Rossetti <i>et al.</i> , (2001)	27/33 (82%) (57/131)* (43%)	ex 8-46 (87%)
<b>PTT &amp; NIRCA</b>	Peral <i>et al.</i> , (1997)	6/95 (6%)	ex 22-46 (42%)
<b>Direct Sequence Analysis</b>	Thomas <i>et al.</i> , (1999)	7/24, (29%)	ex 2-15 (47%)
	Rossetti <i>et al.</i> , (2001)	30/98 (31%) (57/131)* (43%)	ex 1-8,15-17, 22-26 (51%)
	Present study	12/17 (71%)	ex 2-21, 23-34 (71%)
	Present study	4/42 (10%)	44-45 (3%)
	Thongnoppakhun <i>et al.</i> , (2000)	1 patient only	Full-length RNA (100%)

\* Rossetti *et al* (2001) detected 57 novel mutations in 131 patients (exons 1-46). Of these, 27 were detected by the PTT, and 30 by direct sequence analysis.

Obviously a 100% detection rate is to be ultimately desired, however, this is often impractical and it can be technically and financially challenging. In other studies of genes where a large number of private mutations occur, such as the *NI-1* gene associated with neurofibromatosis, detection rates over 50% are considered average across a number of screening methods (Fahsold *et al.*, 2000).

We were unable to detect the probable disease-causing mutations in five of the Australian *PKD1*-linked families screened (A7, A12, A13, A29, A55). The mutation detection study in these families was preceded by linkage analysis, however it should be noted that four of the five families had only a minimum number of four members. A positive lod score was obtained for the chromosome 16 locus and, given the fact that approximately 85% of Caucasian ADPKD cases are due to mutations in *PKD1* (Hughes *et al.*, 1995), it seems highly probable that they are linked to *PKD1* (See Table VII.2). However, with so few related individuals investigated, one cannot rule out the possibility that the low positive lod score was a chance finding.

**Table VII.2.** Observed Lod Score Values for *PKD1* families with undetected mutation

Family (Number of Individuals tested)	Observed Lod Score Maximum	Observed Lod Score Minimum	Probability of <i>PKD1</i> linkage*
A7 (4)	0.2993	0.2984	91%
A12 (4)	0.3009	0.3005	91%
A13 (4)	0.2780	0.2779	91%
A29 (5)	0.5568	0.5566	95%
A55 (4)	0.5809	0.5809	95%

\* Probability of linkage to *PKD1* was calculated as previously described (Hateboer *et al.*, 1999a). If probability was between 86-100% linkage to *PKD1* was inferred.

Assuming linkage to *PKD1*, our failure to detect the mutation may be technical in origin, in other words due to low sensitivity of the detection system. A more likely

possibility is that these families may have mutations in regions of the *PKD1* gene that were not screened. As stated previously, exons 1, 22 and 42 were not analysed, and large areas of the untranslated region, including the promoter region, were also omitted. The presence of gross deletions in the ADPKD allele is another possible reason why disease-causing mutations were not detected. In this case, only the wild-type allele would be amplified for mutation screening. For example, cryptic splice-sites may be activated in an intronic region that had not been screened. This form of variation could lead to the truncation of polycystin-1 but the genomic variation in the screened fragments would remain undetected. To test for this possibility it would be necessary to amplify the full length RNA and examine the size of the resulting transcriptional products.

Analysis of the mutations that were detected in the present study showed that 15 of the 24 possible pathogenic variants (62.5%) resulted from a single base substitution and 9 resulted from small deletions or insertions. The nucleotide substitutions comprised 9 transversions and 6 transitions. Of the 50 neutral polymorphisms detected, 4 were the result of a small insertion or deletion and 46 were single base substitutions. This group sub-divided into 38 transitions and 8 transversions.

Although a greater number of transversions throughout the genome would be anticipated to occur by chance (4 possible transversions compared with 2 possible transitions), in reality this has not been reported. When 337 pairs of human and rodent orthologues were compared, it was found that the transition rate exceeded the transversion rate by a ratio of 1.4 to 1 for missense mutations and greater than 2.0 to 1 for silent polymorphisms (Collins & Jukes, 1994). More recently work on the Human

Genome Project has determined a transition to transversion ratio of 1.89:1 (Venter et al., 2001). A total of 31 transversions and 63 transitions have been reported in the literature as *PKD1* mutations which result from a single base substitution. This represents a transition to transversion ratio of approximately 2.1:1 for pathogenic variants, when neutral variants are considered this ratio changes to of 3.25:1 (see Tables II.3, II.4.a & II.4.b).

When the single base substitutions detected in the present study (both pathogenic and silent) are combined with those reported in the literature, 72% were transitions and 28% transversions. This represents a transition to transversion ratio of 2.6 to 1, more in accordance with that reported by Venter *et al.* (2001).

In classifying DNA variants as possibly pathogenic or non-pathogenic the guidelines proposed by Cotton & Scriver (1998) generally were followed. Mutations were classified as disease-causing when they could be predicted to result in a truncated protein or, in the case of missense mutations, segregated with the disease phenotype and caused a change in the predicted secondary structure of the protein (Rost & Sander, 1994). For the non-truncating mutations, the evolutionary conservation of the affected protein sequence was also examined as another indication of the possible functional importance of the particular codon(s). One hundred normal chromosomes for all mutations in the duplicated region of the gene were not tested as recommended (Cotton & Scriver, 1998), because the demanding nature of very long range amplification of *PKD1* fragments made an extended analysis of this type impractical.



Twenty-four of the DNA variants detected in the study were classified as probable disease-causing mutations. Twenty-one of these variants were novel and three had been published previously (Turco *et al.*, 1995; Rossetti *et al.*, 1996; and Daniells *et al.*, 1998b). Subdivision on the basis of the effect of the mutation on gene expression showed that between 50% to 62.5% were truncating, these included seven nonsense mutations, four small deletions and one large deletion (12/24). A further three splice-site mutations may also result in a truncation (3/24). The remaining 9/24 variants (37.5%) were missense or in-frame deletions or insertions.

The frequency of truncating mutations in the unique region of *PKD1* was 66%, and in the duplicated region of the gene it was 58%. Other studies (listed in Table II.3) have reported that around 70% of mutations in *PKD1* were truncating. These reported values may be slightly biased towards an over-representation of truncating mutations, as a number of studies (Roelfsema *et al.*, 1997; Peral *et al.*, 1997; Rossetti *et al.*, 2001) favoured screening methods such as PTT and NIRCA which are specifically designed to detect truncating mutations. Taken together, the results of *PKD1* mutation detection studies suggest that the exclusive use of techniques aimed at the identification of truncating mutations would miss a significant proportion (around 30%) of the pathogenic variants causing ADPKD.

Nine of the probable disease-causing mutations detected in this study were predicted to result in an intact protein with an altered amino acid sequence. These include five missense mutations, three in-frame deletions and one in-frame insertion. These mutations were examined (Chapter V) in terms of evolutionary conservation and predicted effect on the secondary structure of the protein (Rost & Sander, 1994). If an

amino acid (or group of amino acids) within a protein is conserved between species during evolution, the conserved residue is more likely to be functionally important than a non-conserved amino acid. In addition to conservation during evolution, the predicted effect on protein topology was also analysed using the PHDsec program (Rost & Sander, 1994). This program is based on the combination of evolutionary information and neural networks and has a reported accuracy of over 72% (Rost & Sander, 1994). When the disease-segregating non-truncating variants detected in the *PKD1* gene in this study were analysed using these two measurements, it was inferred that they were more likely to be pathogenic if they altered a conserved residue and significantly changed the predicted topology of the protein.

A further three probable pathogenic mutations were the product of nucleotide substitutions in the canonical splice signals or near the splice site. The effect of these possible splice site mutations has to be determined by RNA studies. Unfortunately, RNA samples were not available at the time of the project. Initial blood samples from Australian index patients were divided equally, DNA was extracted from one half of the sample, and immortalised lymphoblastoid cell lines were created from the other half. Due to equipment malfunction these cell lines were accidentally thawed during the course of the study. Patients will therefore need to be re-bled in the future in order to allow RNA experiments to be performed.

Two of the mutations, IVS14+1 G>C and IVS16+1 G>T, resulted in the replacement of guanine at the GT donor site. This form of sequence alteration infers a loss of the donor site that is likely to result in the activation of a cryptic donor splice-site in the

following intron, or in the continuation of translation up to the beginning of the next exon.

In the simplest prediction, the IVS 14+1 G>C mutation could cause the addition of five novel amino acids prior to the utilisation of a cryptic GT splice donor site at genomic position 26957-5814. Alternatively, if this cryptic splice site is not activated, 14 novel amino acids may be translated before the introduction of a novel stop codon at genomic position 26984-86.

Likewise, the IVS 16+1 G>T alteration may result in the addition of 23 novel amino acids prior to the introduction of a cryptic GT splice donor site at genomic position 31463-64. Alternatively, 25 novel amino acids may be translated prior to the introduction of a novel stop codon at genomic position 31469-71.

The third possible splicing mutation, IVS37-10 C>A, is more complex. The change is sited 10bp before the start of exon 38 in IVS37 and could constitute an acceptor site alteration, which in turn may lead to the activation of a cryptic AG splice acceptor site (see Fig V.3 for the proposed model). If this cryptic splice-site is activated, the resultant frame-shift could cause the addition of 13 residues to the protein prior to the introduction of a novel stop codon. As shown in Chapter V, of the six family members tested three were affected and had this mutation, and the three members who were phenotypically unaffected did not. The change was not detected in 47 other unrelated *PKD1*-linked individuals. On the basis of the family segregation and the possible acceptor site change, we presume that this could be the disease-causing mutation in this family, although only RNA studies could confirm/refute this speculation.

The early mutation detection studies were biased towards the unique region of the *PKDI* gene (Turco *et al.*, 1995; Peral *et al.*, 1995; Rossetti *et al.*, 1996; Neophytou *et al.*, 1996). Screening techniques in the duplicated region that favoured the detection of truncating mutations followed these studies (Roelfsema *et al.*, 1997; Peral *et al.*, 1997). Reports on analyses of the entire *PKDI* gene for all types of mutation have begun to appear in the scientific literature only recently (Thomas *et al.*, 1999; Perrichot *et al.*, 1999, 2000a & b; Rossetti *et al.*, 2001; Phakdeekitcharoen *et al.*, 2000 & 2001). Truncating and in-frame pathogenic mutations have been detected across the *PKDI* gene from exon 1, which corresponds to the flanking region around the leucine-rich repeat domain, to exon 46 which translates into the coiled-coil domain in the cytoplasmic region prior to the carboxy-terminus.

When the type of mutation is considered in the context of its position, trends begin to emerge in the published data that point to the possible existence of mutation hot spots in *PKDI* which deserve special attention in mutation detection studies. It should be noted, however, that as the number of *PKDI* mutations identified increases these trends may prove to be illusions. The location of the novel mutations detected in this study, and of previously reported mutations (see Table II.3), indicate that in the populations tested a high proportion of pathogenic variants cluster in exons 15 to 18, 23 and 25, and in exons 44 to 46. At the protein level, these areas correspond to the PKD repeat motifs, the REJ domain, the final three transmembrane regions, the coiled-coil domain and the cytoplasmic carboxy terminus.

A clustering of truncating mutations appears to exist in two areas of the gene, one within the unique region and another one within the duplicated region. Within the unique region, most truncating mutations are located in the final three exons of the gene, corresponding to the loops in the extracellular and cytoplasmic domains, the final transmembrane domains, the coiled-coil domain and the carboxyl-terminus of polycystin-1. With the exception of exon 42, the entire unique region was analysed in this study. Clustering was observed, with 29% (7/24) of mutations located in exons 44 to 46, which account for only 5.6% of the coding region. The majority (4/24) were nonsense mutations in exon 44. These four mutations constitute 9% of all truncating mutations observed in our study population and are clustered in just 1.1% of the gene.

In the duplicated region, truncating mutations clustered in exon 15, with 24 of the 29 mutations identified so far in this exon resulting in premature truncation of the protein. Only one study (Roelfsema *et al.*, 1997) used PTT to detect five mutations in this exon. The remaining 19 mutations have been identified using a variety of techniques (Watnick *et al.*, 1999; Thomas *et al.*, 1999; Perrichot *et al.*, 2000b; Phakdeekitcharoen *et al.*, 2000; Rossetti *et al.*, 2001;), suggesting that the observed clustering may be real. Thus it appears that exons 15, 44, 45 and 46 may be suitable candidates for PTT screening.

Missense mutations predominate in exons 23 and 25 in the duplicated region of the gene. These two exons harbour 15% (26/168) of all mutations reported to-date (Peral *et al.*, 1997; Watnick *et al.*, 1997; Rossetti *et al.*, 2001; Phakdeekitcharoen *et al.*, 2001), and the majority of these mutations (76%) are missense changes. Exons 23 and 25 are adjacent to the polypyrimidine tracts in introns 21 and 22 (Van Raay *et al.*, 1996). The high GC content of the tracts makes them particularly susceptible to SNP variation,

since methylated cytosines in CpG dinucleotides are susceptible to forming thymine and account for a 10-fold increase in mutation rate when compared to other dinucleotides (Venter *et al.*, 2001). These tracts have been implicated in gene conversion events (Watnick *et al.*, 1998a), with the homologous genes acting as reservoirs for inter-locus gene conversion.

The second area where mutations seem to cluster in the duplicated region corresponds to exons 16, 17 and 18 (Watnick *et al.*, 1999; Perrichot *et al.*, 2000b; Phakdeekitcharoen *et al.*, 2000; Rossetti *et al.*, 2001; Bouba *et al.*, 2001). These exons, encoding the REJ domain of polycystin-1, account for 4.5% of the *PKD1* coding sequence but harbour 7.6% of all reported mutations. No particular type of mutation has been observed to predominate in this region.

Like others to date (Krawczak & Cooper, 1997; Afzal *et al.*, 2000; Aguiari *et al.*, 2000; Kim *et al.*, 2000b; Koptides *et al.*, 2000; Perrichot *et al.*, 2000 a & b; Phakdeekitcharoen *et al.*, 2000 & 2001; Rossetti *et al.*, 2001; Bouba *et al.*, 2001), this study failed to identify pathogenic variants in exons 3, 6, 7, 9, 14, 26, 29, 30 and 33. This may be due to several reasons. Firstly, it could be that these regions of the gene are not essential for protein function, and so mutations would remain undetected due to ascertainment bias towards ADPKD-affected individuals. Only two neutral polymorphisms have been detected in these exons; a novel missense variation in exon 14 (M/T1092) and a previously described silent exonic change (P3110) in exon 26 (Peral *et al.*, 1997). Conversely, pathogenic mutations in these regions may be so detrimental that they are not conducive to post-natal survival and so they would not be found in the adult population. As greater numbers of individuals are screened for *PKD1*

mutations, it will become apparent if these regions are indeed devoid of disease-causing mutations in the adult population.

Most mutation detection studies of the *PKD1* gene have been performed on the DNA of Caucasian individuals, primarily those of northern European descent. Two studies have focussed on the DNA of Asian individuals from Thailand and Korea respectively (Phakdeekitcharoen *et al.*, 2000; Kim, *et al.*, 2000b). The first mutations in exons 18 and 19 of the gene (R2408C, R2430X, c.7535ins3, Q2558X) were reported exclusively in the Thai population (Phakdeekitcharoen *et al.*, 2000). As yet, there are insufficient published data on ethnic background to establish whether ethnic differences do exist in the location and type of pathogenic sequence variation present. For example, three recurrent mutations in exon 44 (Q4010X, R4020X and Q4041X) have been recorded in Anglo-Irish, southern European and Asian families (Daniells *et al.*, 1998b; Kim *et al.*, 2000b; Rossetti *et al.*, 1996; Torra *et al.*, 1998), providing further evidence that exon 44 is a hot spot for truncating mutations. A slipped-mispairing model has been suggested as the mutational mechanism for Q4041X (Daniells *et al.*, 1998b).

An interesting observation on disease-causing mutations in the *PKD1* gene is their private nature, *i.e.* their restriction to single affected families, which account for the large number of novel mutations discovered in mutation detection studies. In this project, 21/24 or 87.5% of the probable pathogenic variants were novel, which is comparable to the 85.2% (69/81 mutations) published recently by Rossetti *et al.* (2001).

The high rate of private mutations in *PKD1* seems to contradict the familial nature of ADPKD, where only 10% of all patients are reported to have *de novo* mutations (Gabow, 1993). One might expect a proportion of the remaining 90% of individuals with ADPKD of familial origin to perhaps share a few common founder mutations. For example, in this study we detected the same mutation (Q4003X) in two Australian patients (A18.1 & A41.1) of Macedonian origin. Although they did not believe themselves to be closely related, they shared common haplotypes and a founder mutation can be presumed. Peral *et al.*, (1995) also reported a similar situation where two families shared the same mutation (E43IVS43 + 5-13del20) and haplotype analysis suggested identity by descent.

The scarcity of founder mutations in the *PKD1* gene may be related to the reduced reproductive fitness of affected individuals. A phenomenon that could theoretically contribute to this effect is that of clinical anticipation in ADPKD (Geberth *et al.*, 1995a; Peral *et al.*, 1996a; Torra *et al.*, 1997; Perrichot *et al.*, 2001), where the age of onset of the disease, and particularly the age at end-stage renal failure, tends to be earlier with each successive generation. If anticipation in ADPKD exists, it may result in declining reproductive fitness in the younger generations of affected families, and thus ultimately lead to the elimination of the familial mutation.

The high mutation rate in *PKD1* could be attributed, at least partly, to the high number of potential gene conversion reservoirs. In addition to the homologous genes located on chromosome 16p13.1, four functional genes have been identified in the paralogous polycystin family, *PKD2*, *PKD2L*, *PKD2L2* and *PKDREJ* (Veldhuisen *et al.*, 1999).



All four paralogues share homology with the *PKD1* gene between exons 38 to 45 in the transmembrane regions and the cytoplasmic tail prior to the coiled-coil domain (Veldhuisen *et al.*, 1999). The high incidence of reported mutations in exons 44 and 45 might therefore be directly correlated with the number of paralogous genes available for gene conversion in this region.

Other mutational mechanisms in the *PKD1* gene have been proposed. As stated previously, a slipped-mispairing model for the Q4041X mutation has been postulated (Daniells *et al.*, 1998b). In that model a directly adjacent imperfect 5 nucleotide repeat exists. If there is slipped mispairing between the nucleotide repeats then T replaces C at nucleotide 12332 and Q4041X results (Daniells *et al.*, 1998b). Other regions of the gene may also be susceptible to such mutational events and give rise to novel variations in the gene. By combining the mutational mechanism theories it is possible that gene conversion, the high GC content of the gene, and the slipped mispairing model may all contribute to the high level of private mutation.

We have detected 50 polymorphic variants, including 31 that have been published previously and 19 novel changes. They represent 42 exonic (13 missense and 29 silent) and eight intronic variations. This should not be interpreted as suggesting the existence of fewer polymorphic variants in the non-coding regions of the gene, but rather reflects an analytical bias since only 50 to 100bp of the intronic sequences surrounding exons were screened in the study.

As of February 2002 the total number of reported non-disease causing polymorphisms in the *PKD1* gene was 132 (Krawczak & Cooper, 1997; Afzal *et al.*, 2000; Aguiari *et*

*et al.*, 2000; Bogdanova *et al.*, 2000; Kim *et al.*, 2000b; Koptides *et al.*, 2000; Perrichot *et al.*, 2000 a & b; Phakdeekitcharoen *et al.*, 2000 & 2001; Rossetti *et al.*, 2001; Mizoguchi *et al.*, 2001; Bouba *et al.*, 2001; Tsuchiya *et al.*, 2001). Of these, 96 are silent variations and 36 are missense variations. The localisation of polymorphisms replicates that of pathogenic mutations in some parts of the gene. For example, 16% (21/131) of polymorphisms occur in exons 23 to 25, around the polypyrimidine tracts where a clustering of pathogenic missense mutations has also been observed, thus providing further support to the gene conversion theory (Watnick *et al.*, 1998a). However, there did not appear to be a clustering of polymorphisms in exons 15 or 44 to 46 which, as discussed earlier in this chapter, harbour a large number of disease-causing truncating mutations.

Particular attention in this study was given to polymorphisms that did not segregate with the disease phenotype, yet cause an amino acid substitution in polycystin-1. Missense changes present a serious interpretational problem in any mutation detection study where a functional assay is unavailable as direct proof of the pathogenic effect. As demonstrated in the study, computer modelling of secondary protein structures predicts changes in a significant proportion of the missense polymorphisms detected. In the case of ADPKD the substantial number of novel mutations compounds the problem, where possible functional effects need to be invoked in every new case. One of the conventional approaches used to determine whether a missense mutation is indeed associated with a given disease phenotype, is to screen 50-100 normal unrelated controls for the mutation and, if present in the controls, the mutation is classified as a non-pathogenic polymorphism. However, besides the prohibitive costs involved in

screening *PKD1* in 50-100 normal controls, the rarity of individual variants in this gene would mean that lack of a mutation in the control subjects could still be inconclusive.

Family segregation may also be insufficient evidence in favour of a pathogenic role. For example, an intronic variation in *PKD1* (IVS41+5 ins 3) was originally reported to be a likely disease-causing change (Perrichot *et al.*, 2000a). It was presumed that the change was involved in aberrant splicing even though RNA analysis was not performed. This insertion mutation segregated with the disease in a French family and was not detected in 100 normal chromosomes, nor had it been reported previously in the literature. Soon after the initial report, the same variant was described as a non-disease causing polymorphism in a large Greek-Cypriot kindred, where the mutation failed to segregate with the ADPKD phenotype (Koptides *et al.*, 2000). Again, however, no RNA analysis was performed in this family. The same insertion mutation was detected by this candidate in Bulgarian patient, B23.1. Subsequent investigation revealed that it segregated with the disease in family B23 (personal communication Dr. Nadja Bogdanova, Münster, Germany). RNA was not available to further investigate if the change had any effect on transcription in family B23. Its role therefore remains to be assessed in future studies.

The so-called “neutral” polymorphisms, especially missense mutations, may have subtle effects on protein function and thus act as modifiers of disease severity in cases where a drastic mutation is the obvious major cause of the disorder. Moreover, the accumulation of such “neutral” polymorphisms may lead to a cooperative effect on protein function, through numerous changes in topology, even in the absence of an obvious pathogenic mutation. The existence of such “hypermorphic” alleles has been

proposed by Watnick *et al.* (1998a). The polymorphisms detected in the present study of five families where no pathogenic mutation was found are summarised in Table VII.3. Four polymorphisms located between exons 2 and 15 of *PKD1* were detected in Australian patient A7.1. Three of the variants were silent in the coding region and one was intronic. No variation was detected in patient A12.3, which was unusual as variation was detected in almost all of the other patient DNA samples analysed in this study. Five common polymorphisms in linkage disequilibrium were detected in patient A55.1 between exons 44 and 46. Ten polymorphisms were detected in patient A13.1 between exons 5 and 18, nine of which were silent variants in the coding region and one was intronic. One patient (A29.1) is a likely candidate for a hypermorphic allele, if such an allele exists. This patient had 15 polymorphisms from exons 13 to 46; of which 6 were missense, 8 were silent changes in the coding region, and 1 was intronic.

**Table VII.3.** Polymorphisms in *PKD1*-linked patients with an undetected mutation

Patient	Polymorphisms	Type of Change	Comments
A7.1	c.487G>A, A92 (1/15) IVS6-68G>A (1/17) c.3322G>A, L1037 (4/17) c.4885G>A, I558T (2/15)	Silent exonic Silent intronic Silent exonic Silent exonic	(1/131) Rossetti <i>et al.</i> , (2001)
A12.3	none detected		
A13.1	c.1330T>C, L373 (5/15) g.23307del7 (3/17) c.2911G>A, P900 (3/17) c.2941C>T, D910 (3/17) c.4876A>C, A1555 (5/15) c.5383C>T, T1724 (5/15) c.5974G>A, L1921 (3/15) c.7138C>T, G2309 (1/15) c.7376T>C, L2389 (4/15) c.7652C>T, L2481 (4/15)	Silent exonic Silent intronic Silent exonic Silent exonic Silent exonic Silent exonic Silent exonic Silent exonic Silent exonic	g.23307del7, P900 and D910 occur in LD in (3/17) patients.
A29.1	c.3274T>C, G1021 (3/17) c.3486T>C, M1092T (3/17) c.4406T>C, W1399R (2/15) c.4876A>C, A1555 (5/15) c.5383C>T, T1724 (5/15) c.7652C>T, L2481 (4/15) c.8124A>G, H2638R (2/15) c.9406GT>CC, F3066L (5/15) c.9541T>C, P3110 (5/15) c.10976C>T, L3589 (2/48) c.10986G>T, S3592I (2/48) IVS44+23insA (2/48) c.12341A>G, I4044V (32/88) c.12484A>G, A4091 (31/85) c.12838T>C, P4209 (31/89)	Silent exonic Missense Missense Silent exonic Silent exonic Silent exonic Missense Missense Silent exonic Silent exonic Missense Silent intronic Missense Silent exonic Silent exonic	L3589 & IVS44+23 in LD I4044V, A4091, P4209, in LD
A55.1	c.12341A>G, I4044V (32/88) c.12384C>T, A4058V (9/51) c.12484A>G, A4091 (31/85) c.12617C>T, L4136 (9/51) c.12838T>C, P4209 (31/89)	Missense Missense Silent exonic Silent exonic Silent exonic	Again I4044V, A4091, P4209, in LD with the subset of A4058V and L4136

**Table Explained:**

Please note that numbers listed in brackets after described polymorphisms represent the number of individuals with the polymorphism from the total number screened, e.g. c.487G>A, A92 (1/15) represents 1 individual with the c.487G>A from the 15 tested.

ADPKD is a well known example of phenotype variability (Milutinovic *et al.*, 1992; Torra *et al.*, 1995; Hateboer *et al.*, 1999a), with marked differences in disease severity observed both between and within families. By analogy to other single gene disorders (eg. neurofibromatosis (*NF1* gene) Fahsold *et al.*, 2000), mutation heterogeneity should be expected to account, at least partly, for the interfamilial variation. Again by analogy with other single gene disorders, such as those described in chapter II eg. phenylketonuria and cystic fibrosis, mutation detection studies should provide information on genotype-phenotype correlations. This will allow conclusions to be drawn on the functional importance of different protein domains, as well as predictions on disease severity in the preclinical testing of at-risk individuals.

In the case of ADPKD, the study of genotype-phenotype correlations is rendered particularly difficult by different factors, some of which are biological in nature, while others are the product of incomplete research or publication design. Biological factors include the private nature of *PKD1* mutations, meaning that the number of individuals carrying the same mutation will never be sufficient for statistically significant conclusions to be made. Another very important factor is related to the proposed two-hit model of cystogenesis (Knudson, 1971; Qian *et al.*, 1996), where the random nature of the second hit could have a major confounding effect on attempts to study the phenotypic consequences of any germline mutation. In addition, the pathogenesis of ADPKD is extremely complicated, with multifactorial traits such as hypertension, known to play a major role in disease severity. This would suggest that a large number of genetic and environmental variables need to be taken into account which, in combination with the rarity of individual mutations and the two-hit model of cystogenesis, makes genotype-phenotype correlations in ADPKD an impossible task.

While the above factors are biological in nature and may thus present insurmountable obstacles, the study of genotype-phenotype correlations in ADPKD is hampered further by the failure of most research groups to provide adequate clinical information in the mutation reports. To date only 16 of the 31 published articles on *PKD1* mutations have accompanying phenotype information, and this is usually limited to the age at onset of end-stage renal failure (ESRF). The general paucity of detail on age at renal failure and associated symptoms could be resolved by increased communication between clinicians and diagnostic laboratories, and the introduction of a comprehensive ADPKD-phenotype database. Such a project was commenced by the PKD Foundation and administered at Johns Hopkins University, USA. Unfortunately, due to lack of funding the database was not maintained and a comparable substitute project has yet to be initiated (personal communication Dr. Gregory Germino, Johns Hopkins University, USA).

Even less information exists on phenotype severity as it relates to the occurrence of supposed neutral sequence variations in the *PKD1* gene. In future, as greater emphasis is placed not only on the search for disease-causing mutations, but also on the variable phenotypes that accompany them, it will be of vital importance to collate these data from the expanding number of private mutations in the gene.

This PhD project was part of a larger collaborative study, resulting in an extensive clinical database containing various measures of phenotype severity, which should allow future analysis of the effect of both pathogenic mutations and polymorphisms on the clinical manifestations of ADPKD.

Here phenotype trends as related to the type of mutation (truncating versus missense) in the patients included in the present study, and in other publications have been examined in detail. Age at onset of ESRF was used as a measure of phenotype severity. In the study, 12 individuals with truncating mutations and 4 with in-frame mutations had reached ESRF. Obviously the number of individuals was too small to attempt to test for significance. Those with a truncating mutation developed ESRF at a mean age of 51.1 years (range = 42-74 years) and those with an in-frame variation developed ESRF at 65.7 years (range = 59-73 years). When truncating versus non-truncating mutations reported in the literature were compared with respect to age at onset of ESRF, 42 individuals with truncating mutations had a mean age at onset of 50.1 years (range = 29-70 years), while 15 individuals with in-frame mutations had a mean age at onset of ESRF of 53.3 years (range = 43-68 years). The difference was not statistically significant in any of the comparisons.

A similar comparison has been conducted by Watnick *et al.* (1999), who found 10 families out of 35 screened to have a clustering of truncating mutations between exons 13 and 17 in the duplicated region of the *PKD1* gene. They proposed that these mutations co-occurred with a specific phenotype, including very early onset of the disease and/or the presence of intracranial aneurysms (Watnick *et al.*, 1999).

In a similar manner the clinical effects of germline mutations in *PKD1* that have been found to occur in more than one family were tested in the present study. The numbers of such mutations are very small and even fewer have been published with phenotype information, see Table VII.4.



**Table VII.4.** Shared pathogenic mutations in *PKD1*

Mutation	Number of families (reference)	Age at ESRF	Other symptoms
c.5224del2	3		
	(Watnick et al., 1999)	N/A	VEO & ICA
	(Watnick et al., 1999)	N/A	ICA
	(Watnick et al., 1999)	N/A	VEO
Q4010X	3		
	(Daniells et al., 1998b)	N/A	N/A
	(Kim et al., 2000b)	N/A	N/A
	Present study	49	Inguinal hernia
R4020X	2		
	(Rossetti et al., 1996)	58,51,49	ICA
	Present study	N/A	VEO
Q4041X	6		
	(Turco et al., 1995)	68,70,42	Anticipation
	(Peral et al., 1997)	N/A	N/A
	(Torra et al., 1998)	42	N/A
	(Daniells et al., 1998b)	N/A	N/A
	(Rossetti et al., 2001)	N/A	N/A
	Present study	42,57	ICA
R4227X	3		
	(Peral et al., 1996b)	42,42	N/A
	(Peral et al., 1997)	57,54,53	N/A
	(Rossetti et al., 2001)	N/A	N/A

N/A – no information available.

VEO – Very early onset of the disease (*in utero* to <30 years).

ICA – Intra-cranial aneurysm.

Given the modest amount of information available at this point in time, it is not possible to ascertain if a direct correlation between the age at onset of ESRF and the same germline mutation exists in different families. However, there may be certain mutations which are associated with the occurrence of intracranial aneurysms and an unusually early age at onset (*in utero* to before 30 years) of cyst development leading to severe renal disease. An ADPKD phenotype with ICA or early onset of the disease with clinical anticipation has been associated with two mutations (R4020X, and Q4041X) in

more than one family, as outlined in Table VII.5. Three unrelated families in the Watnick *et al.* study (1999) shared the same frame-shift mutation in exon 15 of *PKD1*, c.5224del2.

**Table VIII.5.** Private mutations with a severe ADPKD phenotype

Mutation	Phenotype	Reference
IVS14+1G>	ICA	Present study
IVS16+1G>	ICA	Present study
Q3206X	Anticipation. ESRF in generation I age 37, 38 & 39. Generation II ESRF age 26.	Perrichot <i>et al.</i> , 2000b
c.11284insT	Anticipation. ESRF in generation I age 40, generation II normal renal function age 25, generation III palpable cysts at birth.	Perrichot <i>et al.</i> , 1999
S3770X	Anticipation. ESRF in generation I age 70, and in generation II age 42.	Present study
Y3818X	Anticipation. ESRF not reached in affected father age 28 with <i>de novo</i> mutation. Renal cysts detected in one dizygotic twin daughter <i>in utero</i> .	Peral <i>et al.</i> , 1996a
c.12801del28	Anticipation. ESRF in generation I age 50, generation II age 58, generation III ages 60, 58, 53, generation IV chronic renal failure at age 23.	Torra <i>et al.</i> , 1997

## Conclusions and Future Studies

Much work remains to be undertaken on the mutational mechanisms and pathogenic pathways of ADPKD. It is now possible to screen the major disease gene (*PKD1*) for mutations and polymorphisms. Each successive variant detected will contribute to a greater understanding of the *PKD1* gene and its protein products. If phenotypic information is also published (or made publicly available in designated databases), it may be possible in the future to attempt genotype-phenotype correlation studies.

The discovery and further characterisation of additional *PKD1* pseudogenes in this study has led to the amplification of a *PKD1*-specific product for mutation detection screening. This will facilitate future studies that undertake molecular diagnosis of ADPKD linked to the *PKD1* gene.

The differences and similarities between the pseudogene and the *PKD1* sequences will provide a basis for further research at a number of levels. As sequence information increases, the accuracy of *PKD1*-specific primers will improve. By comparing the pseudogene sequences to *PKD1* and each other, the order of evolution of the pseudogenes may also be modelled. *PKD1* pseudogene sequences and paralogous gene sequences from genes in the PKD family will provide evidence to strengthen or dispute the gene conversion theory. As genomic information on other species improves, searches for *PKD1* homologous sequences could be applied to primate genomic libraries in order to assess the stage of molecular evolution at which *PKD1* duplications first occurred.

In addition to providing molecular diagnostic strategies for the *PKD1* gene, this study may benefit other groups investigating genes where co-amplification of homologous sequences presents an obstacle in mutation detection. Similar approaches to those outlined in this study, using BAC libraries, could be used for the unique amplification of other genes, and the study of the function of homologous genes.

For ADPKD families pre-symptomatic diagnosis, preceded by informed consent, can be extremely beneficial for various reasons. For example, individuals who are at-risk can ascertain whether they are likely to develop renal cysts and thus be more likely to develop renal failure with advancing age. Individuals who are pre-symptomatically diagnosed as affected can be monitored, and if necessary treated, for hypertension prior to the onset of renal deterioration. These people then have the opportunity to make informed life-style and reproductive choices. At-risk individuals who are diagnosed as unaffected can be spared unnecessary anxiety about developing the disease and, if they wish, can volunteer as possible kidney donors for affected family members.

As suggested by Prof. Arno Motulsky in his address to the American Society of Human Genetics (2000), in future it will be necessary to understand the pathways that lead from an altered gene sequence to the disease phenotype. A new discipline termed “phenogenetics” is emerging that deals with the study of genotypic expression and how it is influenced by modifying genes, somatic mutations and epigenetic factors (Motulsky, 2000). Given the strong possibility that somatic mutations and modifying factors greatly influence the variable phenotype associated with ADPKD, it will be necessary to understand what influence all of these factors exert on the overall clinical outcome of the disease.

Microarray analysis may provide information of how gene expression levels of other, possible modifying genes, are altered in varying *PKDI* phenotypes both between and within families. Although this analysis would not assist in determining the influence of somatic mutations, it may reduce the number of variables in the overall question. It has been predicted that there are as few as 26,000 to 40,000 protein-coding genes in the human genome (McPherson *et al.*, 2001; Venter *et al.*, 2001). Given the levels of species complexity, this relatively low number of genes highlights the underlying gene interactions as well as the pre- and post-natal environmental conditions which must exist in order to account for phenotypic diversity in humans.

## **Bibliography**

## Bibliography

- Abernathy, C. R., Rasmussen, S. A., Stalker, H. J., Zori, R., Driscoll, D. J., Williams, C. A., Kousseff, B. G., & Wallace, M. R. (1997). NF1 mutation analysis using a combined heteroduplex/SSCP approach. *Hum Mutat*, 9(6), 548-554.
- Afzal, A. R., Florencio, R. N., Taylor, R., Patton, M. A., Saggarr-Malik, A., & Jeffery, S. (2000). Novel mutations in the duplicated region of the polycystic kidney disease 1 (PKD1) gene provides supporting evidence for gene conversion. *Genet Test*, 4(4), 365-370.
- Afzal, A. R., Hand, M., Ternes-Pereira, E., Saggarr-Malik, A., Taylor, R., & Jeffery, S. (1999). Novel mutations in the 3 region of the polycystic kidney disease 1 (PKD1) gene. *Hum Genet*, 105(6), 648-653.
- Afzal, A. R., & Jeffery, S. (2001). Amplification of a 13.5-kb region of the PKD1 gene containing the 2.5-kb polypyrimidine tract in intron 21 facilitates mutation detection in this gene. *Genet Test*, 5(1), 57-59.
- Aguiari, G., Manzati, E., Penolazzi, L., Micheletti, F., Augello, G., Vitali, E. D., Cappelli, G., Cai, Y., Reynolds, D., Somlo, S., Piva, R., & del Senno, L. (1999). Mutations in autosomal dominant polycystic kidney disease 2 gene: Reduced expression of PKD2 protein in lymphoblastoid cells. *Am J Kidney Dis*, 33(5), 880-885.
- Aguiari, G., Savelli, S., Garbo, M., Bozza, A., Augello, G., Penolazzi, L., De Paoli Vitali, E., La Torre, C., Cappelli, G., Piva, R., & del Senno, L. (2000). Novel splicing and missense mutations in autosomal dominant polycystic kidney disease 1 (PKD1) gene: expression of mutated genes. *Hum Mutat*, 16(5), 444-445.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, 215(3), 403-410.
- Andersen, P. M., Nilsson, P., Ala-Hurula, V., Keranen, M. L., Tarvainen, I., Haltia, T., Nilsson, L., Binzer, M., Forsgren, L., & Marklund, S. L. (1995). Amyotrophic lateral sclerosis associated with homozygosity for an Asp90Ala mutation in CuZn-superoxide dismutase. *Nat Genet*, 10(1), 61-66.
- Aoyama, T., Sawamura, T., Furutani, Y., Matsuoka, R., Yoshida, M. C., Fujiwara, H., & Masaki, T. (1999). Structure and chromosomal assignment of the human lectin-like oxidized low-density-lipoprotein receptor-1 (LOX-1) gene. *Biochem J*, 339 (Pt 1), 177-184.
- Aoyama, T., Tynan, K., Dietz, H. C., Francke, U., & Furthmayr, H. (1993). Missense mutations impair intracellular processing of fibrillin and microfibril assembly in Marfan syndrome. *Hum Mol Genet*, 2(12), 2135-2140.

- Ariza, M., Alvarez, V., Marin, R., Aguado, S., Lopez-Larrea, C., Alvarez, J., Menendez, M. J., & Coto, E. (1997). A family with a milder form of adult dominant polycystic kidney disease not linked to the PKD1 (16p) or PKD2 (4q) genes. *J Med Genet*, 34(7), 587-589.
- Arnold, N., Gross, E., Schwarz-Boeger, U., Pfisterer, J., Jonat, W., & Kiechle, M. (1999). A highly sensitive, fast, and economical technique for mutation analysis in hereditary breast and ovarian cancers. *Hum Mutat*, 14(4), 333-339.
- Arnould, T., Sellin, L., Benzing, T., Tsiokas, L., Cohen, H. T., Kim, E., & Walz, G. (1999). Cellular activation triggered by the autosomal dominant polycystic kidney disease gene product PKD2. *Mol Cell Biol*, 19(5), 3423-3434.
- Ausubel, F. M., & Albright, L. M. (1994). Isotopic Sequencing. *Current Protocols in Molecular Biology*, 2(Suppl. 26), Chp 7.
- Baboolal, K., Ravine, D., Daniels, J., Williams, N., Holmans, P., Coles, G. A., & Williams, J. D. (1997). Association of the angiotensin I converting enzyme gene deletion polymorphism with early onset of ESRF in PKD1 adult polycystic kidney disease. *Kidney Int*, 52(3), 607-613.
- Bachner, L., Vinet, M. C., Lacave, R., Babron, M. C., Rondeau, E., Sraer, J. D., Chevet, D., & Kaplan, J. C. (1990). Linkage study of a large family with autosomal dominant polycystic kidney disease with reduced expression. Absence of linkage to the PKD 1 locus. *Hum Genet*, 85(2), 221-227.
- Badenas, C., Torra, R., Darnell, A., & Estivill, X. (1997). Mutations and intragenic polymorphisms in the diagnosis of autosomal dominant polycystic kidney disease type 1. *Contrib Nephrol*, 122, 45-48.
- Badenas, C., Torra, R., Perez-Oller, L., Mallolas, J., Talbot-Wright, R., Torregrosa, V., & Darnell, A. (2000). Loss of heterozygosity in renal and hepatic epithelial cystic cells from ADPKD1 patients. *Eur J Hum Genet*, 8(7), 487-492.
- Badenas, C., Torra, R., San Millan, J. L., Lucero, L., Mila, M., Estivill, X., & Darnell, A. (1999). Mutational analysis within the 3' region of the PKD1 gene. *Kidney Int*, 55(4), 1225-1233.
- Bairoch, A., & Apweiler, R. (1999). The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res*, 27(1), 49-54.
- Bateman, A., & Sandford, R. (1999). The PLAT domain: a new piece in the PKD1 puzzle. *Curr Biol*, 9(16), R588-590.
- Bateman, J. F., Freddi, S., Lamande, S. R., Byers, P., Nasioulas, S., Douglas, J., Otway, R., Kohonen-Corish, M., Ekins, E., & Forrest, S. (1999). Reliable and sensitive detection of premature termination mutations using a protein truncation test designed to overcome problems of nonsense-mediated mRNA instability. *Hum Mutat*, 13(4), 311-317.



- Bear, C. E., Duguay, F., Naismith, A. L., Kartner, N., Hanrahan, J. W., & Riordan, J. R. (1991). Cl<sup>-</sup> channel activity in *Xenopus* oocytes expressing the cystic fibrosis gene. *J Biol Chem*, 266(29), 19142-19145.
- Bear, J. C., Parfrey, P. S., Morgan, J. M., Martin, C. J., & Cramer, B. C. (1992). Autosomal dominant polycystic kidney disease: new information for genetic counselling. *Am J Med Genet*, 43(3), 548-553.
- Bhattacharyya, A., & Lilley, D. M. (1989). The contrasting structures of mismatched DNA sequences containing looped-out bases (bulges) and multiple mismatches (bubbles). *Nucleic Acids Res*, 17(17), 6821-6840.
- Biermann, C. W., Gasser, T. C., Breuer, C., & Rutishauser, G. (1992). [Marfan syndrome and cystic kidneys of the adult type]. *Helv Chir Acta*, 59(3), 513-515.
- Biller, D. S., DiBartola, S. P., Eaton, K. A., Pflueger, S., Wellman, M. L., & Radin, M. J. (1996). Inheritance of polycystic kidney disease in Persian cats. *J Hered*, 87(1), 1-5.
- Blaszyk, H., Hartmann, A., Schroeder, J. J., McGovern, R. M., Sommer, S. S., & Kovach, J. S. (1995). Rapid and efficient screening for p53 gene mutations by dideoxy fingerprinting. *Biotechniques*, 18(2), 256-260.
- Bogdanova, N., Dworniczak, B., Dragova, D., Todorov, V., Dimitrakov, D., Kalinov, K., Hallmayer, J., Horst, J., & Kalaydjieva, L. (1995). Genetic heterogeneity of polycystic kidney disease in Bulgaria. *Hum Genet*, 95(6), 645-650.
- Bogdanova, N., Markoff, A., Gerke, V., McCluskey, M., Horst, J., & Dworniczak, B. (2001). Homologues to the first gene for autosomal dominant polycystic kidney disease are pseudogenes. *Genomics*, 74(3), 333-341.
- Bogdanova, N., McCluskey, M., Sikmann, K., Markoff, A., Todorov, V., Dimitrakov, D., Schiavello, T., Thomas, M., Kalaydjieva, L., Dworniczak, B., & Horst, J. (2000). Screening the 3' region of the polycystic kidney disease 1 (PKD1) gene in 41 Bulgarian and Australian kindreds reveals a prevalence of protein truncating mutations. *Hum Mutat*, 16(2), 166-174.
- Bohn, M., Berge, K. E., Bakken, A., Erikssen, J., & Berg, K. (1993). Insertion/deletion (I/D) polymorphism at the locus for angiotensin I-converting enzyme and myocardial infarction. *Clin Genet*, 44(6), 292-297.
- Bork, P., Holm, L., & Sander, C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol*, 242(4), 309-320.
- Bouba, I., Koptides, M., Mean, R., Costi, C. E., Demetriou, K., Georgiou, I., Pierides, A., Siamopoulos, K., & Deltas, C. C. (2001). Novel PKD1 deletions and missense variants in a cohort of Hellenic polycystic kidney disease families. *Eur J Hum Genet*, 9(9), 677-684.

- Bradbury, N. A. (1999) Intracellular CFTR: localization and function. *Physiol Rev*, 79(1 Suppl), S175-191.
- Brahe, C., & Bertini, E. (1996). Spinal muscular atrophies: recent insights and impact on molecular diagnosis. *J Mol Med*, 74(10), 555-562.
- Brasier, J. L., & Henske, E. P. (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(2), 194-199.
- Breuning, M. H., Snijdwint, F. G., Brunner, H., Verwest, A., Ijdo, J. W., Saris, J. J., Dauwerse, J. G., Blonden, L., Keith, T., Callen, D. F., & et al. (1990). Map of 16 polymorphic loci on the short arm of chromosome 16 close to the polycystic kidney disease gene (PKD1). *J Med Genet*, 27(10), 603-613.
- Brill, S. R., Ross, K. E., Davidow, C. J., Ye, M., Grantham, J. J., & Caplan, M. J. (1996). Immunolocalization of ion transport proteins in human autosomal dominant polycystic kidney epithelial cells. *Proc Natl Acad Sci U S A*, 93(19), 10206-10211.
- Brook-Carter, P. T., Peral, B., Ward, C. J., Thompson, P., Hughes, J., Maheshwar, M. M., Nellist, M., Gamble, V., Harris, P. C., & Sampson, J. R. (1994). Deletion of the TSC2 and PKD1 genes associated with severe infantile polycystic kidney disease—a contiguous gene syndrome. *Nat Genet*, 8(4), 328-332.
- Brookes, A. J., Lehvaslaiho, H., Siegfried, M., Boehm, J. G., Yuan, Y. P., Sarkar, C. M., Bork, P., & Ortigao, F. (2000). HGBASE: a database of SNPs and other variations in and around human genes. *Nucleic Acids Res*, 28(1), 356-360.
- Budowle, B., Chakraborty, R., Giusti, A. M., Eisenberg, A. J., & Allen, R. C. (1991). Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. *Am J Hum Genet*, 48(1), 137-144.
- Bycroft, M., Bateman, A., Clarke, J., Hamill, S. J., Sandford, R., Thomas, R. L., & Chothia, C. (1999). The structure of a PKD domain from polycystin-1: implications for polycystic kidney disease. *Embo J*, 18(2), 297-305.
- Cai, Y., Maeda, Y., Cedzich, A., Torres, V. E., Wu, G., Hayashi, T., Mochizuki, T., Park, J. H., Witzgall, R., & Somlo, S. (1999). Identification and characterization of polycystin-2, the PKD2 gene product. *J Biol Chem*, 274(40), 28557-28565.
- Carone, F. A., Bacallao, R., & Kanwar, Y. S. (1995). The pathogenesis of polycystic kidney disease. *Histol Histopathol*, 10(1), 213-221.
- Caulfield, M., Lavender, P., Farrall, M., Munroe, P., Lawson, M., Turner, P., & Clark, A. J. (1994). Linkage of the angiotensinogen gene to essential hypertension. *N Engl J Med*, 330(23), 1629-1633.

- Cerasola, G., Li Vecchi, M., Mule, G., Cottone, S., Mangano, M. T., Andronico, G., Contomo, A., Parrino, I. A., Renda, F., Pavone, G., & Scialabba, A. (1997). Role of renin-angiotensin-aldosterone system and of sympathetic activity in arterial hypertension associated with autosomal dominant polycystic kidney disease. *Contrib Nephrol*, 122, 22-27.
- Chapman, A. B., Johnson, A., Gabow, P. A., & Schrier, R. W. (1990). The renin-angiotensin-aldosterone system and autosomal dominant polycystic kidney disease. *N Engl J Med*, 323(16), 1091-1096.
- Chapman, A. B., Rubinstein, D., Hughes, R., Stears, J. C., Earnest, M. P., Johnson, A. M., Gabow, P. A., & Kaehny, W. D. (1992). Intracranial aneurysms in autosomal dominant polycystic kidney disease. *N Engl J Med*, 327(13), 916-920.
- Chapman, A. B., & Schrier, R. W. (1991). Pathogenesis of hypertension in autosomal dominant polycystic kidney disease. *Semin Nephrol*, 11(6), 653-660.
- Churchill, D. N., Bear, J. C., Morgan, J., Payne, R. H., McManamon, P. J., & Gault, M. H. (1984). Prognosis of adult onset polycystic kidney disease re-evaluated. *Kidney Int*, 26(2), 190-193.
- Clegg, J. B. (1987). Can the product of the theta gene be a real globin? *Nature*, 329(6138), 465-466.
- Collier, S., Tassabehji, M., Sinnott, P., & Strachan, T. (1993). A de novo pathological point mutation at the 21-hydroxylase locus: implications for gene conversion in the human genome. *Nat Genet*, 3(3), 260-265.
- Collins, D. W., & Jukes, T. H. (1994). Rates of transition and transversion in coding sequences since the human-rodent divergence. *Genomics*, 20(3), 386-396.
- Connacher, A. A., Forsyth, C. C., & Stewart, W. K. (1987). Orofaciodigital syndrome type I associated with polycystic kidneys and agenesis of the corpus callosum. *J Med Genet*, 24(2), 116-118.
- Consortium, T. E. P. K. D. (1994). The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. *Cell*, 77(6), 881-894.
- Constantinides, R., Xenophontos, S., Neophytou, P., Nomura, S., Pierides, A., & Deltas, C. C. (1997). New amino acid polymorphism, Ala/Val4058, in exon 45 of the polycystic kidney disease 1 gene: evolution of alleles. *Hum Genet*, 99(5), 644-647.
- Constantinou-Deltas, C. D., Papageorgiou, E., Boteva, K., Christodoulou, K., Breuning, M. H., Peter, D. J., & Pierides, A. (1995). Genetic heterogeneity in adult dominant polycystic kidney disease in Cypriot families. *Hum Genet*, 95(4), 416-423.

- Constantinou-Deltas, C. D., Papageorgiou, E., Boteva, K., Christodoulou, K., & Pierides, A. (1995). Weak evidence for allelic association in the cypriot PKD1 population. *Contrib Nephrol*, 115, 93-96.
- Correlation between genotype and phenotype in patients with cystic fibrosis. The Cystic Fibrosis Genotype-Phenotype Consortium. (1993). *N Engl J Med*, 329(18), 1308-1313.
- Coto, E., Aguado, S., Alvarez, J., Menendez Diaz, M. J., & Lopez-Larrea, C. (1992). Genetic and clinical studies in autosomal dominant polycystic kidney disease type 1 (ADPKD1). *J Med Genet*, 29(4), 243-246.
- Coto, E., Sanz de Castro, S., Aguado, S., Alvarez, J., Arias, M., Menendez, M. J., & Lopez-Larrea, C. (1995). DNA microsatellite analysis of families with autosomal dominant polycystic kidney disease types 1 and 2: evaluation of clinical heterogeneity between both forms of the disease. *J Med Genet*, 32(6), 442-445.
- Cotton, C. U., & Avner, E. D. (1998). PKD and CF: an interesting family provides insight into the molecular pathophysiology of polycystic kidney disease. *Am J Kidney Dis*, 32(6), 1081-1083.
- Cotton, R. G., Rodrigues, N. R., & Campbell, R. D. (1988). Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc Natl Acad Sci U S A*, 85(12), 4397-4401.
- Cotton, R. G., & Scriver, C. R. (1998). Proof of "disease causing" mutation. *Hum Mutat*, 12(1), 1-3.
- Dalgaard, O. Z. (1957). Bilateral polycystic disease of the kidneys: a follow-up of two hundred and eighty-four patients and their families. *Acta Med Scand Suppl*, 328, 1-255.
- Danaci, M., Akpolat, T., Bastemir, M., Sarikaya, S., Akan, H., Selcuk, M. B., & Cengiz, K. (1998). The prevalence of seminal vesicle cysts in autosomal dominant polycystic kidney disease. *Nephrol Dial Transplant*, 13(11), 2825-2828.
- Daniells, C., Maheshwar, M., Lazarou, L., Davies, F., Coles, G., & Ravine, D. (1998a). Human gene mutations. Gene symbol: PKD1. Disease: Polycystic kidney disease. *Hum Genet*, 102(1), 127.
- Daniells, C., Maheshwar, M., Lazarou, L., Davies, F., Coles, G., & Ravine, D. (1998b). Novel and recurrent mutations in the PKD1 (polycystic kidney disease) gene. *Hum Genet*, 102(2), 216-220.
- Daoust, M. C., Reynolds, D. M., Bichet, D. G., & Somlo, S. (1995). Evidence for a third genetic locus for autosomal dominant polycystic kidney disease. *Genomics*, 25(3), 733-736.

- Davidow, C. J., Maser, R. L., Rome, L. A., Calvet, J. P., & Grantham, J. J. (1996). The cystic fibrosis transmembrane conductance regulator mediates transepithelial fluid secretion by human autosomal dominant polycystic kidney disease epithelium in vitro. *Kidney Int*, 50(1), 208-218.
- Davis, I. D., MacRae Dell, K., Sweeney, W. E., & Avner, E. D. (2001). Can progression of autosomal dominant or autosomal recessive polycystic kidney disease be prevented? *Semin Nephrol*, 21(5), 430-440.
- de Almeida, E., Martins Prata, M., de Almeida, S., & Lavinha, J. (1999). Long-term follow-up of a family with autosomal dominant polycystic kidney disease type 3. *Nephrol Dial Transplant*, 14(3), 531-634.
- de Almeida, S., de Almeida, E., Peters, D., Pinto, J. R., Tavora, I., Lavinha, J., Breuning, M., & Prata, M. M. (1995). Autosomal dominant polycystic kidney disease: evidence for the existence of a third locus in a Portuguese family. *Hum Genet*, 96(1), 83-88.
- de Franchis, R., Kraus, E., Kozich, V., Sebastio, G., & Kraus, J. P. (1999). Four novel mutations in the cystathionine beta-synthase gene: effect of a second linked mutation on the severity of the homocystinuric phenotype. *Hum Mutat*, 13(6), 453-457.
- Devuyst, O. (1999). [Role of chlorine transporters and water channels in autosomal dominant polycystic kidneys]. *Bull Mem Acad R Med Belg*, 154(6 Pt 2), 309-318.
- Di Matteo, J., Picard, R., Vacheron, A., & Bensaid, J. (1965). [Renal polycystosis associated with partial Marfan's syndrome with dilatation of the ascending aorta and aortic insufficiency]. *Bull Mem Soc Med Hop Paris*, 116(16), 1665-1673.
- Ding, L., Zhang, S., Qiu, W., Xiao, C., Wu, S., Zhang, G., & Cheng, L. (2002). Novel mutations of PKD1 gene in Chinese patients with autosomal dominant polycystic kidney disease. *Nephrol Dial Transplant*, 17(1), 75-80.
- Dipple, K. M., & McCabe, E. R. (2000). Phenotypes of patients with "simple" Mendelian disorders are complex traits: thresholds, modifiers, and systems dynamics. *Am J Hum Genet*, 66(6), 1729-1735.
- Dodd, A., Rowland, S. A., Hawkes, S. L., Kennedy, M. A., & Love, D. R. (1997). Mutations in the adrenoleukodystrophy gene. *Hum Mutat*, 9(6), 500-511.
- Donnai, D., Kerzin-Storarr, L., & Harris, R. (1987). Familial orofaciodigital syndrome type I presenting as adult polycystic kidney disease. *J Med Genet*, 24(2), 84-87.
- Doria, A., Warram, J. H., & Krolewski, A. S. (1994). Genetic predisposition to diabetic nephropathy. Evidence for a role of the angiotensin I--converting enzyme gene. *Diabetes*, 43(5), 690-695.

- Drickamer, K. (1989). Multiple subfamilies of carbohydrate recognition domains in animal lectins. *Ciba Found Symp*, 145, 45-58, discussion 58-61.
- Eikenboom, J. C., Vink, T., Briet, E., Sixma, J. J., & Reitsma, P. H. (1994). Multiple substitutions in the von Willebrand factor gene that mimic the pseudogene sequence. *Proc Natl Acad Sci U S A*, 91(6), 2221-2224.
- Elles, R. G., Read, A. P., Hodgkinson, K. A., Watters, A., & Harris, R. (1990). Recombination or heterogeneity: is there a second locus for adult polycystic kidney disease? *J Med Genet*, 27(7), 413-417.
- Evan, A. P., Gardner, K. D., Jr., & Bernstein, J. (1979). Polypoid and papillary epithelial hyperplasia: a potential cause of ductal obstruction in adult polycystic disease. *Kidney Int*, 16(6), 743-750.
- Fahsold, R., Hoffmeyer, S., Mischung, C., Gille, C., Ehlers, C., Kucukceylan, N., Abdel-Nour, M., Gewies, A., Peters, H., Kaufmann, D., Buske, A., Tinschert, S., & Nurnberg, P. (2000). Minor lesion mutational spectrum of the entire NF1 gene does not explain its high mutability but points to a functional domain upstream of the GAP-related domain. *Am J Hum Genet*, 66(3), 790-818.
- Fanen, P., Clain, J., Labarthe, R., Hulin, P., Girodon, E., Pagesy, P., Goossens, M., & Edelman, A. (1999). Structure-function analysis of a double-mutant cystic fibrosis transmembrane conductance regulator protein occurring in disorders related to cystic fibrosis. *FEBS Lett*, 452(3), 371-374.
- Feather, S. A., Woolf, A. S., Donnai, D., Malcolm, S., & Winter, R. M. (1997). The oral-facial-digital syndrome type 1 (OFD1), a cause of polycystic kidney disease and associated malformations, maps to Xp22.2-Xp22.3. *Hum Mol Genet*, 6(7), 1163-1167.
- Feldhahn, J. (1995). Polycystic kidney disease in a persian cat. *Aust Vet Practit*, 25(4), 176-178.
- Ferrari, M., & Cremonesi, L. (1996). Genotype-phenotype correlation in cystic fibrosis patients. *Ann Biol Clin (Paris)*, 54(6), 235-241.
- Fick, G. M., Johnson, A. M., Hammond, W. S., & Gabow, P. A. (1995). Causes of death in autosomal dominant polycystic kidney disease. *J Am Soc Nephrol*, 5(12), 2048-2056.
- Fischer, S. G., & Lerman, L. S. (1979). Length-independent separation of DNA restriction fragments in two-dimensional gel electrophoresis. *Cell*, 16(1), 191-200.
- Fodde, R., & Losekoot, M. (1994). Mutation detection by denaturing gradient gel electrophoresis (DGGE). *Hum Mutat*, 3(2), 83-94.

- Foggensteiner, L., Bevan, A. P., Thomas, R., Coleman, N., Boulter, C., Bradley, J., Ibraghimov-Beskrovnya, O., Klinger, K., & Sandford, R. (2000). Cellular and subcellular distribution of polycystin-2, the protein product of the PKD2 gene. *J Am Soc Nephrol*, 11(5), 814-827.
- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., & Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 151(4), 1531-1545.
- Forrest, S., Cotton, R., Landegren, U., & Southern, E. (1995). How to find all those mutations. *Nat Genet*, 10(4), 375-376.
- Gabow, P. (1996). *Definition and natural history of autosomal dominant polycystic kidney disease* (1st ed.). Oxford: Oxford University Press.
- Gabow, P. A. (1990). Autosomal dominant polycystic kidney disease--more than a renal disease. *Am J Kidney Dis*, 16(5), 403-413.
- Gabow, P. A. (1993). Autosomal dominant polycystic kidney disease. *Am J Kidney Dis*, 22(4), 511-512.
- Gabow, P. A., Johnson, A. M., Kaehny, W. D., Kimberling, W. J., Lezotte, D. C., Duley, I. T., & Jones, R. H. (1992). Factors affecting the progression of renal disease in autosomal-dominant polycystic kidney disease. *Kidney Int*, 41(5), 1311-1319.
- Ganguly, A., Rock, M. J., & Prockop, D. J. (1993). Conformation-sensitive gel electrophoresis for rapid detection of single-base differences in double-stranded PCR products and DNA fragments: evidence for solvent-induced bends in DNA heteroduplexes. *Proc Natl Acad Sci U S A*, 90(21), 10325-10329.
- Geberth, S., Ritz, E., Zeier, M., & Stier, E. (1995). Anticipation of age at renal death in autosomal dominant polycystic kidney disease (ADPKD)? *Nephrol Dial Transplant*, 10(9), 1603-1606.
- Geberth, S., Stier, E., Zeier, M., Mayer, G., Rambauser, M., & Ritz, E. (1995). More adverse renal prognosis of autosomal dominant polycystic kidney disease in families with primary hypertension. *J Am Soc Nephrol*, 6(6), 1643-1648.
- Geng, L., Segal, Y., Peissel, B., Deng, N., Pei, Y., Carone, F., Rennke, H. G., Glucksmann-Kuis, A. M., Schneider, M. C., Ericsson, M., Reeders, S. T., & Zhou, J. (1996). Identification and localization of polycystin, the PKD1 gene product. *J Clin Invest*, 98(12), 2674-2682.
- Germino, G. G., Barton, N. J., Lamb, J., Higgs, D. R., Harris, P., Xiao, G. H., Scherer, G., Nakamura, Y., & Reeders, S. T. (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(5), 925-933.

- Germiño, G. G., Somlo, S., Weinstat-Saslow, D., & Reeders, S. T. (1993). Positional cloning approach to the dominant polycystic kidney disease gene, PKD1. *Kidney Int Suppl*, 39, S20-25.
- Goldrick, M. M., Kimball, G. R., Liu, Q., Martin, L. A., Sommer, S. S., & Tseng, J. Y. (1996). NIRCA: a rapid robust method for screening for unknown point mutations. *Biotechniques*, 21(1), 106-112.
- Gorlin, R. J. (1995). Nevroid basal cell carcinoma syndrome. *Dermatol Clin*, 13(1), 113-125.
- Gouldesbrough, D. R., & Fleming, S. (1998). Unilateral and segmental localised polycystic kidney disease. *J Clin Pathol*, 51(9), 703-705.
- Grantham, J. J. (1997). Pathogenesis of autosomal dominant polycystic kidney disease: recent developments. *Contrib Nephrol*, 122, 1-9.
- Grantham, J. J., Geiser, J. L., & Evan, A. P. (1987). Cyst formation and growth in autosomal dominant polycystic kidney disease. *Kidney Int*, 31(5), 1145-1152.
- Grantham, J. J., Ye, M., Davidow, C., Holub, B., & Sharma, M. (1995). Evidence for a potent lipid secretagogue in the cyst fluids of patients with autosomal dominant polycystic kidney disease. *J Am Soc Nephrol*, 6(4), 1242-1249.
- Griffin, M. D., Torres, V. E., Grande, J. P., & Kumar, R. (1996). Immunolocalization of polycystin in human tissues and cultured cells. *Proc Assoc Am Physicians*, 108(3), 185-197.
- Grunfeld, J. P. (1998). Factors influencing progression of renal failure in autosomal dominant polycystic kidney disease. *Am J Kidney Dis*, 32(5), xlvii-xlviii.
- Grunfeld, J. P., Albouze, G., Jungers, P., Landais, P., Dana, A., Droz, D., Moynot, A., Lafforgue, B., Boursztyń, E., & Franco, D. (1985). Liver changes and complications in adult polycystic kidney disease. *Adv Nephrol Necker Hosp*, 14, 1-20.
- Guillaume, R., D'Agati, V., Daoust, M., & Trudel, M. (1999). Murine Pkd1 is a developmentally regulated gene from morula to adulthood: role in tissue condensation and patterning. *Dev Dyn*, 214(4), 337-348.
- Guo, L., Chen, M., Basora, N., & Zhou, J. (2000). The human polycystic kidney disease 2-like (PKDL) gene: exon/intron structure and evidence for a novel splicing mechanism. *Mamm Genome*, 11(1), 46-50.
- Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P. C., Ottina, K., Wallace, M. R., Sakaguchi, A. Y., & et al. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature*, 306(5940), 234-238.



- Gyllenstein, U. B., & Erlich, H. A. (1988). Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proc Natl Acad Sci US A*, 85(20), 7652-7656.
- Haavik, J., Nishino, H., Liu, Q., & Sommer, S. S. (1996). Bi-directional dideoxy fingerprinting (Bi-ddF): rapid and efficient screening for mutations in the Big Blue transgenic mouse mutation detection system. *Biotechniques*, 20(6), 988-990, 992-984.
- Hanaoka, K., Devuyst, O., Schwiebert, E. M., Wilson, P. D., & Guggino, W. B. (1996). A role for CFTR in human autosomal dominant polycystic kidney disease. *Am J Physiol*, 270(1 Pt 1), C389-399.
- Hanaoka, K., & Guggino, W. B. (2000). cAMP regulates cell proliferation and cyst formation in autosomal polycystic kidney disease cells. *J Am Soc Nephrol*, 11(7), 1179-1187.
- Hao, E., Menke, J. B., Smith, A. M., Jones, C., Geffner, M. E., Hershman, J. M., Wuerth, J. P., Samuels, H. H., Ways, D. K., & Usala, S. J. (1994). Divergent dimerization properties of mutant beta 1 thyroid hormone receptors are associated with different dominant negative activities. *Mol Endocrinol*, 8(7), 841-851.
- Harris, P. C. (1997). The TSC2/PKD1 contiguous gene syndrome. *Contrib Nephrol*, 122, 76-82.
- Harris, P. C. (1999). Autosomal dominant polycystic kidney disease: clues to pathogenesis. *Hum Mol Genet*, 8(10), 1861-1866.
- Harris, P. C., Barton, N. J., Higgs, D. R., Reeders, S. T., & Wilkie, A. O. (1990). A long-range restriction map between the alpha-globin complex and a marker closely linked to the polycystic kidney disease 1 (PKD1) locus. *Genomics*, 7(2), 195-206.
- Harris, P. C., Thomas, S., Ratcliffe, P. J., Breuning, M. H., Coto, E., & Lopez-Larrea, C. (1991). Rapid genetic analysis of families with polycystic kidney disease 1 by means of a microsatellite marker. *Lancet*, 338(8781), 1484-1487.
- Harrod, M. J., Stokes, J., Peede, L. F., & Goldstein, J. L. (1976). Polycystic kidney disease in a patient with the oral-facial-digital syndrome - type I. *Clin Genet*, 9(2), 183-186.
- Hashimoto, C., Hudson, K. L., & Anderson, K. V. (1988). The Toll gene of *Drosophila*, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. *Cell*, 52(2), 269-279.
- Hateboer, N., Buchalter, M., Davies, S. J., Lazarou, L. P., & Ravine, D. (2000). Co-occurrence of autosomal dominant polycystic kidney disease and Marfan syndrome in a kindred. *Am J Kidney Dis*, 35(4), 753-760.

- Hateboer, N., Lazarou, L. P., Williams, A. J., Holmans, P., & Ravine, D. (1999). Familial phenotype differences in PKD11. *Kidney Int*, 56(1), 34-40.
- Hateboer, N., v Dijk, M. A., Bogdanova, N., Coto, E., Saggarr-Malik, A. K., San Millan, J. L., Torra, R., Breuning, M., & Ravine, D. (1999). Comparison of phenotypes of polycystic kidney disease types 1 and 2. European PKD1-PKD2 Study Group. *Lancet*, 353(9147), 103-107.
- Hayashi, K., & Yandell, D. W. (1993). How sensitive is PCR-SSCP? *Hum Mutat*, 2(5), 338-346.
- Hayashi, T., Mochizuki, T., Reynolds, D. M., Wu, G., Cai, Y., & Somlo, S. (1997). Characterization of the exon structure of the polycystic kidney disease 2 gene (PKD2). *Genomics*, 44(1), 131-136.
- Hayashi, Y., Yamamoto, M., Ohmori, S., Kamijo, T., Ogawa, M., & Seo, H. (1999). Inhibition of growth hormone (GH) secretion by a mutant GH-I gene product in neuroendocrine cells containing secretory granules: an implication for isolated GH deficiency inherited in an autosomal dominant manner. *J Clin Endocrinol Metab*, 84(6), 2134-2139.
- Hayward-Lester, A., Oefner, P. J., & Doris, P. A. (1996). Rapid quantification of gene expression by competitive RT-PCR and ion-pair reversed-phase HPLC. *Biotechniques*, 20(2), 250-257.
- Higashihara, E., Nutahara, K., Kojima, M., Tamakoshi, A., Yoshiyuki, O., Sakai, H., & Kurokawa, K. (1998). Prevalence and renal prognosis of diagnosed autosomal dominant polycystic kidney disease in Japan. *Nephron*, 80(4), 421-427.
- Hossack, K. F., Leddy, C. L., Johnson, A. M., Schrier, R. W., & Gabow, P. A. (1988). Echocardiographic findings in autosomal dominant polycystic kidney disease. *N Engl J Med*, 319(14), 907-912.
- Huan, Y., & van Adelsberg, J. (1999). Polycystin-1, the PKD1 gene product, is in a complex containing E-cadherin and the catenins. *J Clin Invest*, 104(10), 1459-1468.
- Hughes, J., Ward, C. J., Aspinwall, R., Butler, R., & Harris, P. C. (1999). Identification of a human homologue of the sea urchin receptor for egg jelly: a polycystic kidney disease-like protein. *Hum Mol Genet*, 8(3), 543-549.
- Hughes, J., Ward, C. J., Peral, B., Aspinwall, R., Clark, K., San Millan, J. L., Gamble, V., & Harris, P. C. (1995). The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. *Nat Genet*, 10(2), 151-160.
- Hulsebos, T. J., Bijleveld, E. H., Riegman, P. H., Smink, L. J., & Dunham, I. (1996). Identification and characterization of NF1-related loci on human chromosomes 22, 14 and 2. *Hum Genet*, 98(1), 7-11.

- Hung, T., Mak, K., & Fong, K. (1990). A specificity enhancer for polymerase chain reaction. *Nucleic Acids Res*, 18(16), 4953.
- Ibraghimov-Beskrovnaya, O., Dackowski, W. R., Foggensteiner, L., Coleman, N., Thiru, S., Petry, L. R., Burn, T. C., Connors, T. D., Van Raay, T., Bradley, J., Qian, F., Onuchic, L. F., Watnick, T. J., Piontek, K., Hakim, R. M., Landes, G. M., Gernino, G. G., Sandford, R., & Klinger, K. W. (1997). Polycystin: in vitro synthesis, in vivo tissue expression, and subcellular localization identifies a large membrane-associated protein. *Proc Natl Acad Sci U S A*, 94(12), 6397-6402.
- Ingnas, M., Byding, S., Eckersten, A., Eriksson, S., Hultman, T., Jorsback, A., Lofman, E., Sabounchi, F., Kressner, U., Lindmark, G., & Tooke, N. (2000). Enzymatic mutation detection in the P53 gene. *Clin Chem*, 46(10), 1562-1573.
- Ioannou, P. A., Amemiya, C. T., Garnes, J., Kroisel, P. M., Shizuya, H., Chen, C., Batzer, M. A., & de Jong, P. J. (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat Genet*, 6(1), 84-89.
- Jordanova, A., Kalaydjieva, L., Savov, A., Claustres, M., Schwarz, M., Estivill, X., Angelicheva, D., Haworth, A., Casals, T., & Kremensky, I. (1997). SSCP analysis: a blind sensitivity trial. *Hum Mutat*, 10(1), 65-70.
- Jordon, C., Harpaz, N., & Thung, S. N. (1989). Caroli's disease and adult polycystic kidney disease: a rarely recognized association. *Liver*, 9, 30-35.
- Kaehny, W. D., & Everson, G. T. (1991). Extrarenal manifestations of autosomal dominant polycystic kidney disease. *Semin Nephrol*, 11(6), 661-670.
- Kaplan, B. S., Kaplan, P., & Kessler, A. (1997). Cystic kidneys associated with connective tissue disorders. *Am J Med Genet*, 69(2), 133-137.
- Kim, K., Drummond, I., Ibraghimov-Beskrovnaya, O., Klinger, K., & Arnaout, M. A. (2000). Polycystin 1 is required for the structural integrity of blood vessels. *Proc Natl Acad Sci U S A*, 97(4), 1731-1736.
- Kim, U. K., Jin, D. K., Ahn, C., Shin, J. H., Lee, K. B., Kim, S. H., Chae, J. J., Hwang, D. Y., Lee, J. G., Namkoong, Y., & Lee, C. C. (2000). Novel mutations of the PKD1 gene in Korean patients with autosomal dominant polycystic kidney disease. *Mutat Res*, 432(1-2), 39-45.
- Kimberling, W. J., Fain, P. R., Kenyon, J. B., Goldgar, D., Sujansky, E., & Gabow, P. A. (1988). Linkage heterogeneity of autosomal dominant polycystic kidney disease. *N Engl J Med*, 319(14), 913-918.
- Kimberling, W. J., Kumar, S., Gabow, P. A., Kenyon, J. B., Connolly, C. J., & Somlo, S. (1993). Autosomal dominant polycystic kidney disease: localization of the second gene to chromosome 4q13-q23. *Genomics*, 18(3), 467-472.
- Kirubakaran, M. G. (1998). The Central Australian Aboriginal Renal Disease Registry. *Nephrology*, 4(Suppl.), S83-S85.

- Klahr, S., Breyer, J. A., Beck, G. J., Dennis, V. W., Hartman, J. A., Roth, D., Steinman, T. I., Wang, S. R., & Yamamoto, M. E. (1995). Dietary protein restriction, blood pressure control, and the progression of polycystic kidney disease. Modification of Diet in Renal Disease Study Group. *J Am Soc Nephrol*, 5(12), 2037-2047.
- Kleff, S., & Kemper, B. (1988). Initiation of heteroduplex-loop repair by T4-encoded endonuclease VII in vitro. *Embo J*, 7(5), 1527-1535.
- Knudson, A. G., Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*, 68(4), 820-823.
- Kohwi, Y., & Panchenko, Y. (1993) Transcription-dependent recombination induced by triple-helix formation. *Genes Dev*, 7(9), 1766-1778.
- Koptides, M., Constantinides, R., Kyriakides, G., Hadjigavriel, M., Patsalis, P. C., Pierides, A., & Deltas, C. C. (1998). Loss of heterozygosity in polycystic kidney disease with a missense mutation in the repeated region of PKD1. *Hum Genet*, 103(6), 709-717.
- Koptides, M., Hadjimichael, C., Koupepidou, P., Pierides, A., & Constantinou Deltas, C. (1999). Germinal and somatic mutations in the PKD2 gene of renal cysts in autosomal dominant polycystic kidney disease. *Hum Mol Genet*, 8(3), 509-513.
- Koptides, M., Mean, R., Demetriou, K., Constantinides, R., Pierides, A., Harris, P. C., & Deltas, C. C. (2000). Screening of the PKD1 duplicated region reveals multiple single nucleotide polymorphisms and a de novo mutation in Hellenic polycystic kidney disease families. *Hum Mutat*, 16(2), 176.
- Kramer, M. F., & Coen, D. M. (1999). The polymerase chain reaction: Enzymatic amplification of DNA by PCR: Standard procedures and optimization. *Current Protocols in Molecular Biology*, 2(Suppl. 46), 15.11.11-15.11.15.
- Krawczak, M., & Cooper, D. N. (1997). The human gene mutation database. *Trends Genet*, 13(3), 121-122.
- Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet*, 22(2), 139-144.
- Kukita, Y., Tahira, T., Sommer, S. S., & Hayashi, K. (1997). SSCP analysis of long DNA fragments in low pH gel. *Hum Mutat*, 10(5), 400-407.
- Kwok, P. Y., Carlson, C., Yager, T. D., Ankener, W., & Nickerson, D. A. (1994). Comparative analysis of human DNA variations by fluorescence-based sequencing of PCR products. *Genomics*, 23(1), 138-144.
- Larsen, L. A., Christiansen, M., Vuust, J., & Andersen, P. S. (1999). High-throughput single-strand conformation polymorphism analysis by automated capillary electrophoresis: robust multiplex analysis and pattern-based identification of allelic variants. *Hum Mutat*, 13(4), 318-327.

- Lazarou, L. P., Davies, F., Sarfarazi, M., Coles, G. A., & Harper, P. S. (1987). Adult polycystic kidney disease and linked RFLPs at the alpha globin locus: a genetic study in the South Wales population. *J Med Genet*, 24(8), 466-473.
- Ledley, F. D. (1991). Clinical application of genotypic diagnosis for phenylketonuria: theoretical considerations. *Eur J Pediatr*, 150(11), 752-756.
- Lee, D. C., Chan, K. W., & Chan, S. Y. (1998). Expression of transforming growth factor alpha and epidermal growth factor receptor in adult polycystic kidney disease. *J Urol*, 159(1), 291-296.
- Lee, J. G., Lee, K. B., Kim, U. K., Ahn, C., Hwang, D. Y., Hwang, Y. H., Eo, H. S., Lee, E. J., Kim, Y. S., Han, J. S., Kim, S., & Lee, J. S. (2001). Genetic heterogeneity in Korean families with autosomal-dominant polycystic kidney disease (ADPKD): the first Asian report. *Clin Genet*, 60(2), 138-144.
- Lerman, L. S., & Silverstein, K. (1987). Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. *Methods Enzymol*, 155, 482-501.
- Levy-Toledano, R., Caro, L. H., Accili, D., & Taylor, S. I. (1994). Investigation of the mechanism of the dominant negative effect of mutations in the tyrosine kinase domain of the insulin receptor. *Embo J*, 13(4), 835-842.
- Li, H. P., Geng, L., Burrow, C. R., & Wilson, P. D. (1999). Identification of phosphorylation sites in the PKD1-encoded protein C-terminal domain. *Biochem Biophys Res Commun*, 259(2), 356-363.
- Liu, Q., Feng, J., & Sommer, S. S. (1996). Bi-directional dideoxy fingerprinting (Bi-ddF): a rapid method for quantitative detection of mutations in genomic regions of 300-600 bp. *Hum Mol Genet*, 5(1), 107-114.
- Liu, W., Smith, D. I., Rechtzigel, K. J., Thibodeau, S. N., & James, C. D. (1998). Denaturing high performance liquid chromatography (DHPLC) used in the detection of germline and somatic mutations. *Nucleic Acids Res*, 26(6), 1396-1400.
- Liu, Y. C., Huang, T. S., Huang, W. K., Chen, C. S., & Tu, H. Z. (1998). Dideoxy fingerprinting for rapid screening of rpoB gene mutations in clinical isolates of Mycobacterium tuberculosis. *J Formos Med Assoc*, 97(6), 400-404.
- Loftus, B. J., Kim, U. J., Sneddon, V. P., Kalush, F., Brandon, R., Fuhrmann, J., Mason, T., Crosby, M. L., Barnstead, M., Cronin, L., Deslattes Mays, A., Cao, Y., Xu, R. X., Kang, H. L., Mitchell, S., Eichler, E. E., Harris, P. C., Venter, J. C., & Adams, M. D. (1999). Genome duplications and other features in 12 Mb of DNA sequence from human chromosome 16p and 16q. *Genomics*, 60(3), 295-308.

- Longa, L., Scolari, F., Brusco, A., Carbonara, C., Polidoro, S., Valzorio, B., Riegler, P., Migone, N., & Maiorca, R. (1997). A large TSC2 and PKD1 gene deletion is associated with renal and extrarenal signs of autosomal dominant polycystic kidney disease. *Nephrol Dial Transplant*, 12(9), 1900-1907.
- Lopez, J. A., Chung, D. W., Fujikawa, K., Hagen, F. S., Papayannopoulou, T., & Roth, G. J. (1987). Cloning of the alpha chain of human platelet glycoprotein Ib: a transmembrane protein with homology to leucine-rich alpha 2-glycoprotein. *Proc Natl Acad Sci U S A*, 84(16), 5615-5619.
- Lozano, A. M., & Leblanc, R. (1992). Cerebral aneurysms and polycystic kidney disease: a critical review. *Can J Neurol Sci*, 19(2), 222-227.
- Lu, W., Peissel, B., Babakhanlou, H., Pavlova, A., Geng, L., Fan, X., Larson, C., Brent, G., & Zhou, J. (1997). Perinatal lethality with kidney and pancreas defects in mice with a targeted Pkd1 mutation. *Nat Genet*, 17(2), 179-181.
- MacDermot, K. D., Saggat-Malik, A. K., Economides, D. L., & Jeffery, S. (1998). Prenatal diagnosis of autosomal dominant polycystic kidney disease (PKD1) presenting in utero and prognosis for very early onset disease. *J Med Genet*, 35(1), 13-16.
- Mallamaci, F., Zuccala, A., Zoccali, C., Testa, A., Gaggi, R., Spoto, B., Martorano, C., Curatola, A., Misefari, V., Cuzzola, F., Romeo, G., & Zucchelli, P. (2000). The deletion polymorphism of the angiotensin-converting enzyme is associated with nephroangiosclerosis. *Am J Hypertens*, 13(4 Pt 1), 433-437.
- Maroni, B. J. (1998). Protein restriction in the pre-end-stage renal disease (ESRD) patient: who, when, how, and the effect on subsequent ESRD outcome. *J Am Soc Nephrol*, 9(12 Suppl), S100-106.
- Martin-Zanca, D., Mitra, G., Long, L. K., & Barbacid, M. (1986). Molecular characterization of the human trk oncogene. *Cold Spring Harb Symp Quant Biol*, 51 Pt 2, 983-992.
- Massie, R. J., Olsen, M., Glaszner, J., Robertson, C. F., & Francis, I. (2000). Newborn screening for cystic fibrosis in Victoria: 10 years' experience (1989-1999). *Med J Aust*, 172(12), 584-587.

McPherson, J. D., Marra, M., Hillier, L., Waterston, R. H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E. R., Wilson, R. K., Fulton, R., Kucaba, T. A., Wagner-McPherson, C., Barbazuk, W. B., Gregory, S. G., Humphray, S. J., French, L., Evans, R. S., Bethel, G., Whittaker, A., Holden, J. L., McCann, O. T., Dunham, A., Soderlund, C., Scott, C. E., Bentley, D. R., Schuler, G., Chen, H. C., Jang, W., Green, E. D., Idol, J. R., Maduro, V. V., Montgomery, K. T., Lee, E., Miller, A., Emerling, S., Kucherlapati, Gibbs, R., Scherer, S., Gorrell, J. H., Sodergren, E., Clerc-Blankenburg, K., Tabor, P., Naylor, S., Garcia, D., de Jong, P. J., Catanese, J. J., Nowak, N., Osoegawa, K., Qin, S., Rowen, L., Madan, A., Dors, M., Hood, L., Trask, B., Friedman, C., Massa, H., Cheung, V. G., Kirsch, I. R., Reid, T., Yonescu, R., Weissenbach, J., Bruls, T., Heilig, R., Branscomb, E., Olsen, A., Doggett, N., Cheng, J. F., Hawkins, T., Myers, R. M., Shang, J., Ramirez, L., Schmutz, J., Velasquez, O., Dixon, K., Stone, N. E., Cox, D. R., Haussler, D., Kent, W. J., Furey, T., Rogic, S., Kennedy, S., Jones, S., Rosenthal, A., Wen, G., Schilhabel, M., Gloeckner, G., Nyakatura, G., Siebert, R., Schlegelberger, B., Korenberg, J., Chen, X. N., Fujiyama, A., Hattori, M., Toyoda, A., Yada, T., Park, H. S., Sakaki, Y., Shimizu, N., Asakawa, S., Kawasaki, K., Sasaki, T., Shintani, A., Shimizu, A., Shibuya, K., Kudoh, J., Minoshima, S., Ramser, J., Seranski, P., Hoff, C., Poustka, A., Reinhardt, R., & Lehrach, H. (2001). A physical map of the human genome. *Nature*, 409(6822), 934-941.

Mechler, B. M. (1987). Isolation of messenger RNA from membrane-bound polysomes. *Methods Enzymol*, 152, 241-248.

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*, 16(3), 1215.

Milutinovic, J. (1989). Massive growth of kidneys in patients with autosomal dominant polycystic kidney disease treated with chronic hemodialysis. *Am J Kidney Dis*, 14(5), 365-368.

Milutinovic, J., Agodoa, L. C., Cutler, R. E., & Striker, G. E. (1980). Autosomal dominant polycystic kidney disease. Early diagnosis and consideration of pathogenesis. *Am J Clin Pathol*, 73(6), 740-747.

Milutinovic, J., Fialkow, P. J., Agodoa, L. Y., Phillips, L. A., Rudd, T. G., & Bryant, J. I. (1984). Autosomal dominant polycystic kidney disease: symptoms and clinical findings. *Q J Med*, 53(212), 511-522.

Milutinovic, J., Rust, P. F., Fialkow, P. J., Agodoa, L. Y., Phillips, L. A., Rudd, T. G., & Sutherland, S. (1992). Intrafamilial phenotypic expression of autosomal dominant polycystic kidney disease. *Am J Kidney Dis*, 19(5), 465-472.

Mizoguchi, M., Tamura, T., Yamaki, A., Higashihara, E., & Shimizu, Y. (2001). Mutations of the PKD1 gene among Japanese autosomal dominant polycystic kidney disease patients, including one heterozygous mutation identified in members of the same family. *J Hum Genet*, 46(9), 511-517.

- Mizoguchi, M., Tamura, T., Yamaki, A., Higashihara, E., & Shimizu, Y. (2002). Genotypes of autosomal dominant polycystic kidney disease in Japanese. *J Hum Genet*, 47(1), 51-54.
- Mochizuki, T., Wu, G., Hayashi, T., Xenophontos, S. L., Veldhuisen, B., Saris, J. J., Reynolds, D. M., Cai, Y., Gabow, P. A., Pierides, A., Kimberling, W. J., Breuning, M. H., Deltas, C. C., Peters, D. J., & Somlo, S. (1996). PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science*, 272(5266), 1339-1342.
- Monaco, A. P., & Larin, Z. (1994). YACs, BACs, PACs and MACs: artificial chromosomes as research tools. *Trends Biotechnol*, 12(7), 280-286.
- Motulsky, A. G. (2000). 1999 ASHG Award for Excellence in Education. Some future directions in medical genetics. *Am J Hum Genet*, 66(4), 1190-1191.
- Mousson, C., Rabec, M., Cercueil, J. P., Virot, J. S., Hillon, P., & Rifflé, G. (1997). Caroli's disease and autosomal dominant polycystic kidney disease: a rare association? *Nephrol Dial Transplant*, 12(7), 1481-1483.
- Moy, G. W., Mendoza, L. M., Schulz, J. R., Swanson, W. J., Glabe, C. G., & Vacquier, V. D. (1996). The sea urchin sperm receptor for egg jelly is a modular protein with extensive homology to the human polycystic kidney disease protein, PKD1. *J Cell Biol*, 133(4), 809-817.
- Murcia, N. S., Sweeney, W. E., Jr., & Avner, E. D. (1999). New insights into the molecular pathophysiology of polycystic kidney disease. *Kidney Int*, 55(4), 1187-1197.
- Murray, V. (1989). Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res*, 17(21), 8889.
- Murti, J. R., Bumbulis, M., & Schimenti, J. C. (1994). Gene conversion between unlinked sequences in the germline of mice. *Genetics*, 137(3), 837-843.
- Myers, R. M., Larin, Z., & Maniatis, T. (1985). Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science*, 230(4731), 1242-1246.
- Nagamine, C. M., Chan, K., & Lau, Y. F. (1989). A PCR artifact: generation of heteroduplexes. *Am J Hum Genet*, 45(2), 337-339.
- Nakai, K., Itoh, C., Miura, Y., Hotta, K., Musha, T., Itoh, T., Miyakawa, T., Iwasaki, R., & Hiramori, K. (1994). Deletion polymorphism of the angiotensin I-converting enzyme gene is associated with serum ACE concentration and increased risk for CAD in the Japanese. *Circulation*, 90(5), 2199-2202.
- Nataraj, A. J., Olivos-Glander, I., Kusakawa, N., & Highsmith, W. E., Jr. (1999). Single-strand conformation polymorphism and heteroduplex analysis for gel-based mutation detection. *Electrophoresis*, 20(6), 1177-1185.



- Neame, P. J., Tapp, H., & Grimm, D. R. (1999). The cartilage-derived, C-type lectin (CLECSF1): structure of the gene and chromosomal location. *Biochim Biophys Acta*, 1446(3), 193-202.
- Nelis, E., Timmerman, V., De Jonghe, P., Van Broeckhoven, C., & Rautenstrauss, B. (1999). Molecular genetics and biology of inherited peripheral neuropathies: a fast-moving field. *Neurogenetics*, 2(3), 137-148.
- Neophytou, P., Constantinides, R., Lazarou, A., Pierides, A., & Deltas, C. C. (1996). Detection of a novel nonsense mutation and an intragenic polymorphism in the PKD1 gene of a Cypriot family with autosomal dominant polycystic kidney disease. *Hum Genet*, 98(4), 437-442.
- Nomura, H., Turco, A. E., Pei, Y., Kalaydjieva, L., Schiavello, T., Weremowicz, S., Ji, W., Morton, C. C., Meisler, M., Reeders, S. T., & Zhou, J. (1998). Identification of PKDL, a novel polycystic kidney disease 2-like gene whose murine homologue is deleted in mice with kidney and retinal defects. *J Biol Chem*, 273(40), 25967-25973.
- Norton, I. D., Pokorny, C. S., Painter, D. M., Johnson, J. R., & Perkins, K. W. (1995). Fraternal sisters with adult polycystic kidney disease and adenoma of the ampulla of Vater. *Gastroenterology*, 109(6), 2007-2010.
- Nousia-Arvanitakis, S. (1999). Cystic fibrosis and the pancreas: recent scientific advances. *J Clin Gastroenterol*, 29(2), 138-142.
- O'Leary, C. A., Mackay, B. M., Malik, R., Edmondston, J. E., Robinson, W. F., & Huxtable, C. R. (1999). Polycystic kidney disease in bull terriers: an autosomal dominant inherited disorder. *Aust Vet J*, 77(6), 361-366.
- Olsson, P. G., Lohning, C., Horsley, S., Kearney, L., Harris, P. C., & Frischauf, A. (1996). The mouse homologue of the polycystic kidney disease gene (Pkd1) is a single-copy gene. *Genomics*, 34(2), 233-235.
- Ong, A. C. (1999). Cyst formation in ADPKD: new insights from natural and targeted mutants. *Nephrol Dial Transplant*, 14(3), 544-546.
- Ong, A. C., Ward, C. J., Butler, R. J., Biddolph, S., Bowker, C., Torra, R., Pei, Y., & Harris, P. C. (1999). Coordinate expression of the autosomal dominant polycystic kidney disease proteins, polycystin-2 and polycystin-1, in normal and cystic tissue. *Am J Pathol*, 154(6), 1721-1729.
- Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., & Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci U S A*, 86(8), 2766-2770.
- Ortino, O., Bonanni, F., Ruffino, C., Maiolino, L., & Tedoldi, A. (1988). [Hepato-renal polycystosis, Marfan's syndrome and spina bifida occulta: a complex association. Description of a clinical case]. *Minerva Med*, 79(12), 1105-1107.

- O'Sullivan, D. A., Torres, V. E., Gabow, P. A., Thibodeau, S. N., King, B. F., & Bergstralh, E. J. (1998). Cystic fibrosis and the phenotypic expression of autosomal dominant polycystic kidney disease. *Am J Kidney Dis*, 32(6), 976-983.
- Palsson, R., Sharma, C. P., Kim, K., McLaughlin, M., Brown, D., & Arnaout, M. A. (1996). Characterization and cell distribution of polycystin, the product of autosomal dominant polycystic kidney disease gene 1. *Mol Med*, 2(6), 702-711.
- Parfrey, P. S., Bear, J. C., Morgan, J., Cramer, B. C., McManamon, P. J., Gault, M. H., Churchill, D. N., Singh, M., Hewitt, R., Somlo, S., & et al. (1990). The diagnosis and prognosis of autosomal dominant polycystic kidney disease. *N Engl J Med*, 323(16), 1085-1090.
- Paterson, A. D., & Pei, Y. (1998). Is there a third gene for autosomal dominant polycystic kidney disease? *Kidney Int*, 54(5), 1759-1761.
- Pearson, W. R., & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A*, 85(8), 2444-2448.
- Pei, Y., Paterson, A. D., Wang, K. R., He, N., Hefferton, D., Watnick, T., Germino, G. G., Parfrey, P., Somlo, S., & St George-Hyslop, P. (2001). Bilineal disease and trans-heterozygotes in autosomal dominant polycystic kidney disease. *Am J Hum Genet*, 68(2), 355-363.
- Pennekamp, P., Bogdanova, N., Wilda, M., Markoff, A., Hameister, H., Horst, J., & Dworniczak, B. (1998). Characterization of the murine polycystic kidney disease (Pkd2) gene. *Mamm Genome*, 9(9), 749-752.
- Peral, B., Gamble, V., San Millan, J. L., Strong, C., Sloane-Stanley, J., Moreno, F., & Harris, P. C. (1995). Splicing mutations of the polycystic kidney disease 1 (PKD1) gene induced by intronic deletion. *Hum Mol Genet*, 4(4), 569-574.
- Peral, B., Gamble, V., Strong, C., Ong, A. C., Sloane-Stanley, J., Zerres, K., Winearls, C. G., & Harris, P. C. (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(6), 1399-1410.
- Peral, B., Ong, A. C., San Millan, J. L., Gamble, V., Rees, L., & Harris, P. C. (1996). A stable, nonsense mutation associated with a case of infantile onset polycystic kidney disease 1 (PKD1). *Hum Mol Genet*, 5(4), 539-542.
- Peral, B., San Millan, J. L., Ong, A. C., Gamble, V., Ward, C. J., Strong, C., & Harris, P. C. (1996). Screening the 3' region of the polycystic kidney disease 1 (PKD1) gene reveals six novel mutations. *Am J Hum Genet*, 58(1), 86-96.
- Peral, B., Ward, C. J., San Millan, J. L., Thomas, S., Stallings, R. L., Moreno, F., & Harris, P. C. (1994). Evidence of linkage disequilibrium in the Spanish polycystic kidney disease I population. *Am J Hum Genet*, 54(5), 899-908.

- Perez-Oller, L., Torra, R., Badenas, C., Mila, M., & Darnell, A. (1999). Influence of the ACE gene polymorphism in the progression of renal failure in autosomal dominant polycystic kidney disease. *Am J Kidney Dis*, 34(2), 273-278.
- Perrichot, R., Mercier, B., Carre, A., Cledes, J., & Ferec, C. (2000). Identification of 3 novel mutations (Y4236X, Q3820X, 11745+2 ins3) in autosomal dominant polycystic kidney disease 1 gene (PKD1). *Hum Mutat*, 15(6), 582.
- Perrichot, R., Mercier, B., Quere, I., Carre, A., Simon, P., Whebe, B., Cledes, J., & Ferec, C. (2000). Novel mutations in the duplicated region of PKD1 gene. *Eur J Hum Genet*, 8(5), 353-359.
- Perrichot, R. A., Mercier, B., de Parscau, L., Simon, P. M., Cledes, J., & Ferec, C. (2001). Inheritance of a stable mutation in a family with early-onset disease. *Nephron*, 87(4), 340-345.
- Perrichot, R. A., Mercier, B., Simon, P. M., Whebe, B., Cledes, J., & Ferec, C. (1999). DGGE screening of PKD1 gene reveals novel mutations in a large cohort of 146 unrelated patients. *Hum Genet*, 105(3), 231-239.
- Persu, A., & Devuyst, O. (2000). Transepithelial chloride secretion and cystogenesis in autosomal dominant polycystic kidney disease. *Nephrol Dial Transplant*, 15(6), 747-750.
- Persu, A., Devuyst, O., Lannoy, N., Mamer, R., Brosnahan, G., Gabow, P. A., Pirson, Y., & Verellen-Dumoulin, C. (2000). CF gene and cystic fibrosis transmembrane conductance regulator expression in autosomal dominant polycystic kidney disease. *J Am Soc Nephrol*, 11(12), 2285-2296.
- Peters, D. J., & Sandkuijl, L. A. (1992). Genetic heterogeneity of polycystic kidney disease in Europe. *Contrib Nephrol*, 97, 128-139.
- Peters, D. J., Spruit, L., Klingel, R., Prins, F., Baelde, H. J., Giordano, P. C., Bernini, L. F., de Heer, E., Breuning, M. H., & Bruijn, J. A. (1996). Adult, fetal, and polycystic kidney expression of polycystin, the polycystic kidney disease-1 gene product. *Lab Invest*, 75(2), 221-230.
- Peters, D. J., Spruit, L., Saris, J. J., Ravine, D., Sandkuijl, L. A., Fossdal, R., Boersma, J., van Eijk, R., Norby, S., Constantinou-Deltas, C. D., & et al. (1993). Chromosome 4 localization of a second gene for autosomal dominant polycystic kidney disease. *Nat Genet*, 5(4), 359-362.
- Peters, D. J., van de Wal, A., Spruit, L., Saris, J. J., Breuning, M. H., Bruijn, J. A., & de Heer, E. (1999). Cellular localization and tissue distribution of polycystin-1. *J Pathol*, 188(4), 439-446.
- Phakdeekitcharoen, B., Watnick, T. J., Ahn, C., Whang, D. Y., Burkhart, B., & Germino, G. G. (2000). Thirteen novel mutations of the replicated region of PKD1 in an Asian population. *Kidney Int*, 58(4), 1400-1412.

- Phakdeekitcharoen, B., Watnick, T. J., & Germino, G. G. (2001). Mutation analysis of the entire replicated portion of PKD1 using genomic DNA samples. *J Am Soc Nephrol*, 12(5), 955-963.
- Piontek, K. B., & Germino, G. G. (1999). Murine Pkd1 introns 21 and 22 lack the extreme polypyrimidine bias present in human PKD1. *Mamm Genome*, 10(2), 194-196.
- Polycystic kidney disease: the complete structure of the PKD1 gene and its protein. (1995). *Cell*, 81(2), 289-298.
- Pound, S. E., Carothers, A. D., Pignatelli, P. M., Macnicol, A. M., Watson, M. L., & Wright, A. F. (1992). Evidence for linkage disequilibrium between D16S94 and the adult onset polycystic kidney disease (PKD1) gene. *J Med Genet*, 29(4), 247-248.
- Prasad, N., O'Kane, K. P., Johnstone, H. A., Wheeldon, N. M., McMahon, A. D., Webb, D. J., & MacDonald, T. M. (1994). The relationship between blood pressure and left ventricular mass in essential hypertension is observed only in the presence of the angiotensin-converting enzyme gene deletion allele. *Qjm*, 87(11), 659-662.
- Presneau, N., Laplace-Marieze, V., Sylvain, V., Lortholary, A., Hardouin, A., Bernard-Gallon, D., & Bignon, Y. J. (1998). New mechanism of BRCA-1 mutation by deletion/insertion at the same nucleotide position in three unrelated French breast/ovarian cancer families. *Hum Genet*, 103(3), 334-339.
- Qian, F., Germino, F. J., Cai, Y., Zhang, X., Somlo, S., & Germino, G. G. (1997). PKD1 interacts with PKD2 through a probable coiled-coil domain. *Nat Genet*, 16(2), 179-183.
- Qian, F., Watnick, T. J., Onuchic, L. F., & Germino, G. G. (1996). The molecular basis of focal cyst formation in human autosomal dominant polycystic kidney disease type I. *Cell*, 87(6), 979-987.
- Ramirez, F., Gayraud, B., & Pereira, L. (1999). Marfan syndrome: new clues to genotype-phenotype correlations. *Ann Med*, 31(3), 202-207.
- Ravine, D., Gibson, R. N., Walker, R. G., Sheffield, L. J., Kincaid-Smith, P., & Danks, D. M. (1994). Evaluation of ultrasonographic diagnostic criteria for autosomal dominant polycystic kidney disease 1. *Lancet*, 343(8901), 824-827.
- Ravine, D., Walker, R. G., Gibson, R. N., Forrest, S. M., Richards, R. I., Friend, K., Sheffield, L. J., Kincaid-Smith, P., & Danks, D. M. (1992). Phenotype and genotype heterogeneity in autosomal dominant polycystic kidney disease. *Lancet*, 340(8831), 1330-1333.

- Reeders, S. T., Breuning, M. H., Davies, K. E., Nicholls, R. D., Jarman, A. P., Higgs, D. R., Pearson, P. L., & Weatherall, D. J. (1985). A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature*, 317(6037), 542-544.
- Reeders, S. T., Breuning, M. H., Ryyanen, M. A., Wright, A. F., Davies, K. E., King, A. W., Watson, M. L., & Weatherall, D. J. (1987). A study of genetic linkage heterogeneity in adult polycystic kidney disease. *Hum Genet*, 76(4), 348-351.
- Reeders, S. T., Keith, T., Green, P., Germino, G. G., Barton, N. J., Lehmann, O. J., Brown, V. A., Phipps, P., Morgan, J., Bear, J. C., & et al. (1988). Regional localization of the autosomal dominant polycystic kidney disease locus. *Genomics*, 3(2), 150-155.
- Reinke, R., Krantz, D. E., Yen, D., & Zipursky, S. L. (1988). Chaoptin, a cell surface glycoprotein required for Drosophila photoreceptor cell morphogenesis, contains a repeat motif found in yeast and human. *Cell*, 52(2), 291-301.
- Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J. L., & et al. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, 245(4922), 1066-1073.
- Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, 273(5281), 1516-1517.
- Roelfsema, J. H., Spruit, L., Saris, J. J., Chang, P., Pirson, Y., van Ommen, G. J., Peters, D. J., & Breuning, M. H. (1997). Mutation detection in the repeated part of the PKD1 gene. *Am J Hum Genet*, 61(5), 1044-1052.
- Roest, P. A., Roberts, R. G., Sugino, S., van Ommen, G. J., & den Dunnen, J. T. (1993). Protein truncation test (PTT) for rapid detection of translation-terminating mutations. *Hum Mol Genet*, 2(10), 1719-1721.
- Roses, A. D., & Saunders, A. M. (1997). Apolipoprotein E genotyping as a diagnostic adjunct for Alzheimer's disease. *Int Psychogeriatr*, 9 Suppl 1, 277-288; discussion 317-221.
- Rossetti, S., Bresin, E., Restagno, G., Carbonara, A., Corra, S., De Prisco, O., Pignatti, P. F., & Turco, A. E. (1996). Autosomal dominant polycystic kidney disease (ADPKD) in an Italian family carrying a novel nonsense mutation and two missense changes in exons 44 and 45 of the PKD1 Gene. *Am J Med Genet*, 65(2), 155-159.
- Rossetti, S., Englisch, S., Bresin, E., Pignatti, P. F., & Turco, A. E. (1997). Detection of mutations in human genes by a new rapid method: cleavage fragment length polymorphism analysis (CFLPA). *Mol Cell Probes*, 11(2), 155-160.

- Rossetti, S., Surmecki, L., Gamble, V., Burton, S., Sneddon, V., Peral, B., Koy, S., Bakkaloglu, A., Komel, R., Winearls, C. G., & Harris, P. C. (2001). Mutation analysis of the entire PKD1 gene: genetic and diagnostic implications. *Am J Hum Genet*, 68(1), 46-63.
- Rost, B., & Sander, C. (1994). Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins*, 19(1), 55-72.
- Rost, B., Sander, C., & Schneider, R. (1994). PHD--an automatic mail server for protein secondary structure prediction. *Comput Appl Biosci*, 10(1), 53-60.
- Rothberg, J. M., Jacobs, J. R., Goodman, C. S., & Artavanis-Tsakonas, S. (1990). slit: an extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains. *Genes Dev*, 4(12A), 2169-2187.
- Rozmahel, R., Wilschanski, M., Matin, A., Plyte, S., Oliver, M., Auerbach, W., Moore, A., Forstner, J., Durie, P., Nadeau, J., Bear, C., & Tsui, L. C. (1996). Modulation of disease severity in cystic fibrosis transmembrane conductance regulator deficient mice by a secondary genetic factor. *Nat Genet*, 12(3), 280-287.
- Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1986). Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature*, 324(6093), 163-166.
- Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular Cloning: a laboratory manual* (2nd ed.). Cold Spring Harbour, N.Y.: Cold Spring Harbour Laboratory.
- Sandford, R., Sgotto, B., Aparicio, S., Brenner, S., Vaudin, M., Wilson, R. K., Chisoe, S., Pepin, K., Bateman, A., Chothia, C., Hughes, J., & Harris, P. (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(9), 1483-1489.
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J., & Roe, B. A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J Mol Biol*, 143(2), 161-178.
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, 74(12), 5463-5467.
- Sarkar, G., Yoon, H. S., & Sommer, S. S. (1992). Dideoxy fingerprinting (ddE): a rapid and efficient screen for the presence of mutations. *Genomics*, 13(2), 441-443.
- Savov, A., Angelicheva, D., Balassopoulou, A., Jordanova, A., Noussia-Arvanitakis, S., & Kalaydjieva, L. (1995). Double mutant alleles: are they rare? *Hum Mol Genet*, 4(7), 1169-1171.

- Scheff, R. T., Zuckerman, G., Harter, H., Delmez, J., & Koehler, R. (1980). Diverticular disease in patients with chronic renal failure due to polycystic kidney disease. *Ann Intern Med*, 92(2 Pt 1), 202-204.
- Schiavello, T., Burke, V., Bogdanova, N., Jasik, P., Melsom, S., Boudville, N., Robertson, K., Angelicheva, D., Dworniczak, B., Lemmens, M., Horst, J., Todorov, V., Dimitrakov, D., Sulowicz, W., Krasniak, A., Stompor, T., Beilin, L., Hallmayer, J., Kalaydjieva, L., & Thomas, M. (2001). Angiotensin-converting enzyme activity and the ACE Alu polymorphism in autosomal dominant polycystic kidney disease. *Nephrol Dial Transplant*, 16(12), 2323-2327.
- Schievink, W. I., Torres, V. E., Piepgras, D. G., & Wiebers, D. O. (1992). Saccular intracranial aneurysms in autosomal dominant polycystic kidney disease. *J Am Soc Nephrol*, 3(1), 88-95.
- Schunkert, H., Hense, H. W., Holmer, S. R., Stender, M., Perz, S., Keil, U., Lorell, B. H., & Riegger, G. A. (1994). Association between a deletion polymorphism of the angiotensin-converting-enzyme gene and left ventricular hypertrophy. *N Engl J Med*, 330(23), 1634-1638.
- Scriber, C. R. (1995). Whatever happened to PKU? *Clin Biochem*, 28(2), 137-144.
- Scriber, C. R., Byck, S., Prevost, L., & Hoang, L. (1996). The phenylalanine hydroxylase locus: a marker for the history of phenylketonuria and human genetic diversity. PAH Mutation Analysis Consortium. *Ciba Found Symp*, 197, 73-90; discussion 90-76.
- Scriber, C. R., & Waters, P. J. (1999). Monogenic traits are not simple: lessons from phenylketonuria. *Trends Genet*, 15(7), 267-272.
- Selgas, R., Temes, J. L., Sobrino, J. A., Viguer, J. M., Otero, A., & Sanchez Sicilia, L. (1981). [Polycystic renal disease in the adult associated with an incomplete form of Marfan's syndrome (author's transl)]. *Med Clin (Barc)*, 76(7), 311-313.
- Serafini, F. M., & Carey, L. C. (1999). Adenoma of the ampulla of Vater: a genetic condition? *HPB Surg*, 11(3), 191-193.
- Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y., & Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc Natl Acad Sci U S A*, 89(18), 8794-8797.
- Slade, M. J., Kirby, R. B., Poci, I., Jones, J. K., & Price, R. G. (1998). Presence of laminin fragments in cyst fluid from patients with autosomal dominant polycystic kidney disease (ADPKD): role in proliferation of tubular epithelial cells. *Biochim Biophys Acta*, 1401(2), 203-210.

- Smith, K. D., Kemp, S., Braiterman, L. T., Lu, J. F., Wei, H. M., Geraghty, M., Stetten, G., Bergin, J. S., Pevsner, J., & Watkins, P. A. (1999). X-linked adrenoleukodystrophy: genes, mutations, and phenotypes. *Neurochem Res*, 24(4), 521-535.
- Somlo, S., Rutecki, G., Giuffra, L. A., Reeders, S. T., Cugino, A., & Whittier, F. C. (1993). A kindred exhibiting cosegregation of an overlap connective tissue disorder and the chromosome 16 linked form of autosomal dominant polycystic kidney disease. *J Am Soc Nephrol*, 4(6), 1371-1378.
- Somlo, S., Wirth, B., Germino, G. G., Weinstat-Saslow, D., Gillespie, G. A., Himmelbauer, H., Steevens, L., Coucke, P., Willems, P., Bachner, L., & et al. (1992). Fine genetic localization of the gene for autosomal dominant polycystic kidney disease (PKD1) with respect to physically mapped markers. *Genomics*, 13(1), 152-158.
- Starnes, H. F., Jr., Lazarus, J. M., & Vineyard, G. (1985). Surgery for diverticulitis in renal failure. *Dis Colon Rectum*, 28(11), 827-831.
- Steffann, J., Vidaud, D., Bousquet, S., Jullien, M., Ninot, A., Kaplan, J. C., Beldjord, C., & Biennu, T. (1998). Novel double mutant CF allele identified in a cystic fibrosis patient with meconium ileus. *Ann Genet*, 41(4), 213-215.
- Strachan, T., & Read, A. P. (1999). *Human molecular genetics* (2nd ed.). Oxford: BIOS Scientific Publishers Ltd.
- Tabor, S., & Richardson, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc Natl Acad Sci U S A*, 84(14), 4767-4771.
- Tan, P., Briner, J., Boltshauser, E., Davis, M. R., Wilton, S. D., North, K., Wallgren-Pettersson, C., & Laing, N. G. (1999). Homozygosity for a nonsense mutation in the alpha-tropomyosin slow gene TPM3 in a patient with severe infantile nemaline myopathy. *Neuromuscul Disord*, 9(8), 573-579.
- Taylor, G. R., & Deeble, J. (1999). Enzymatic methods for mutation scanning. *Genet Anal*, 14(5-6), 181-186.
- Theophilus, B. D., Latham, T., Grabowski, G. A., & Smith, F. I. (1989). Comparison of RNase A, a chemical cleavage and GC-clamped denaturing gradient gel electrophoresis for the detection of mutations in exon 9 of the human acid beta-glucosidase gene. *Nucleic Acids Res*, 17(19), 7707-7722.
- Thomas, R., McConnell, R., Whittaker, 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(1), 39-49.



- Thongnoppakhun, A., Rungroj, N., Wilairat, P., Vareesangthip, K., Sirinavin, C., & Yenchitsomanus, P. T. (2000). A novel splice-acceptor site mutation (IVS13-2A>T) of polycystic kidney disease 1 (PKD1) gene resulting in an RNA processing defect with a 74-nucleotide deletion in exon 14 of the mRNA transcript. *Hum Mutat*, 15(1), 115.
- Thongnoppakhun, W., Wilairat, P., Vareesangthip, K., & Yenchitsomanus, P. T. (1999). Long RT-PCR Amplification of the entire coding sequence of the polycystic kidney disease 1 (PKD1) gene. *Biotechniques*, 26(1), 126-132.
- Tighe, O., Carmody, M., & Croke, D. T. (1998). Identification of a novel PKD1 mutation in an Irish autosomal dominant polycystic kidney disease kindred. *Biochem Soc Trans*, 26(3), S265.
- Torra, R., Badenas, C., Darnell, A., Bru, C., Escorsell, A., & Estivill, X. (1997). Autosomal dominant polycystic kidney disease with anticipation and Caroli's disease associated with a PKD1 mutation. Rapid communication. *Kidney Int*, 52(1), 33-38.
- Torra, R., Badenas, C., Peral, B., Darnell, A., Serra, E., Gamble, V., Turco, A. E., Harris, P. C., & Estivill, X. (1998). Recurrence of the PKD1 nonsense mutation Q4041X in Spanish, Italian, and British families. *Hum Mutat, Suppl 1*, S117-120.
- Torra, R., Badenas, C., San Millan, J. L., Perez-Oller, L., Estivill, X., & Darnell, A. (1999). A loss-of-function model for cystogenesis in human autosomal dominant polycystic kidney disease type 2. *Am J Hum Genet*, 65(2), 345-352.
- Torra, R., Darnell, A., Estivill, X., Botey, A., & Revert, L. (1995). Interfamilial and intrafamilial variability of clinical expression in ADPKD. *Contrib Nephrol*, 115, 97-101.
- Tsiokas, L., Kim, E., Arnould, T., Sukhatme, V. P., & Walz, G. (1997). Homo- and heterodimeric interactions between the gene products of PKD1 and PKD2. *Proc Natl Acad Sci U S A*, 94(13), 6965-6970.
- Tsuchiya, K., Komeda, M., Takahashi, M., Yamashita, N., Cigira, M., Suzuki, T., Suzuki, K., Nihei, H., & Mochizuki, T. (2001). Mutationai analysis within the 3' region of the PKD1 gene in Japanese families. *Mutat Res*, 458(3-4), 77-84.
- Tuma, R. S., Beaudet, M. P., Jin, X., Jones, L. J., Cheung, C. Y., Yue, S., & Singer, V. L. (1999). Characterization of SYBR Gold nucleic acid gel stain: a dye optimized for use with 300-nm ultraviolet transilluminators. *Anal Biochem*, 268(2), 278-288.

- Turco, A. E., Bresin, E., Rossetti, S., Englisch, S., Pignatti, P. F., Gammara, L., Maschio, G., Bendetti, M., Li Vecchi, M., Ferrantelli, A., Cerasola, G., Stiasny, B., & Schulze, B. (1997). Molecular genetic investigations in autosomal dominant polycystic kidney disease. Gene Mutation detection, linkage analysis, and preliminary ACE gene I/D polymorphism association studies: an update. *Contrib Nephrol*, 122, 53-57.
- Turco, A. E., Clementi, M., Rossetti, S., Tenconi, R., & Pignatti, P. F. (1996). An Italian family with autosomal dominant polycystic kidney disease unlinked to either the PKD1 or PKD2 gene. *Am J Kidney Dis*, 28(5), 759-761.
- Turco, A. E., Padovani, E. M., Chiaffoni, G. P., Peissel, B., Rossetti, S., Marcolongo, A., Gammara, L., Maschio, G., & Pignatti, P. F. (1993). Molecular genetic diagnosis of autosomal dominant polycystic kidney disease in a newborn with bilateral cystic kidneys detected prenatally and multiple skeletal malformations. *J Med Genet*, 30(5), 419-422.
- Turco, A. E., Rossetti, S., Bresin, E., Corra, S., Gammara, L., Maschio, G., & Pignatti, P. F. (1995). A novel nonsense mutation in the PKD1 gene (C3817T) is associated with autosomal dominant polycystic kidney disease (ADPKD) in a large three-generation Italian family. *Hum Mol Genet*, 4(8), 1331-1335.
- Turco, A. E., Rossetti, S., Bresin, E., Englisch, S., Corra, S., & Pignatti, P. F. (1997). Three novel mutations of the PKD1 gene in Italian families with autosomal dominant polycystic kidney disease. *Hum Mutat*, 10(2), 164-167.
- Tusie-Luna, M. T., & White, P. C. (1995). Gene conversions and unequal crossovers between CYP21 (steroid 21-hydroxylase gene) and CYP21P involve different mechanisms. *Proc Natl Acad Sci U S A*, 92(23), 10796-10800.
- Vacquier, V. D., & Moy, G. W. (1997). The fucose sulfate polymer of egg jelly binds to sperm REJ and is the inducer of the sea urchin sperm acrosome reaction. *Dev Biol*, 192(1), 125-135.
- van Adelsberg, J. (1999). Peptides from the PKD repeats of polycystin, the PKD1 gene product, modulate pattern formation in the developing kidney. *Dev Genet*, 24(3-4), 299-308.
- Van Adelsberg, J., Chamberlain, S., & D'Agati, V. (1997). Polycystin expression is temporally and spatially regulated during renal development. *Am J Physiol*, 272(5 Pt 2), F602-609.
- van Dijk, M. A., Chang, P. C., Peters, D. J., & Breuning, M. H. (1995). Intracranial aneurysms in polycystic kidney disease linked to chromosome 4. *J Am Soc Nephrol*, 6(6), 1670-1673.
- van Dijk, M. A., Peters, D. J., Breuning, M. H., & Chang, P. C. (1999). The angiotensin-converting enzyme genotype and microalbuminuria in autosomal dominant polycystic kidney disease. *J Am Soc Nephrol*, 10(9), 1916-1920.

- Van Raay, T. J., Burn, T. C., Connors, T. D., Petry, L. R., Germino, G. G., Klinger, K. W., & Landes, G. M. (1996). A 2.5 kb polypyrimidine tract in the PKD1 gene contains at least 23 H-DNA-forming sequences. *Microb Comp Genomics*, 1(4), 317-327.
- Veldhuisen, B., Spruit, L., Dauwerse, H. G., Breuning, M. H., & Peters, D. J. (1999). Genes homologous to the autosomal dominant polycystic kidney disease genes (PKD1 and PKD2). *Eur J Hum Genet*, 7(8), 860-872.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yoosheph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X.,

- Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., & Zhu, X. (2001). The sequence of the human genome. *Science*, 291(5507), 1304-1351.
- Wakabayashi, T., Fujita, S., Ohbora, Y., Suyama, T., Tamaki, N., & Matsumoto, S. (1983). Polycystic kidney disease and intracranial aneurysms. Early angiographic diagnosis and early operation for the unruptured aneurysm. *J Neurosurg*, 58(4), 488-491.
- Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., Lander, E. S., & et al. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, 280(5366), 1077-1082.
- Wang, G., Seidman, M. M., & Glazer, P. M. (1996). Mutagenesis in mammalian cells induced by triple helix formation and transcription-coupled repair. *Science*, 271(5250), 802-805.
- Wang, H., Kuwata, S., Juji, T., Yanagisawa, M., Tokunaga, K., Horie, S., Higashihara, E., Kurokawa, K., Yoshikura, H., & Shibata, Y. (1995). Ethnic differences in allele frequencies of two microsatellite markers closely linked to the locus for polycystic kidney disease 1 (PKD1). *Hum Hered*, 45(2), 84-89.
- Wang, M., Kishnani, P., Decker-Phillips, M., Kahler, S. G., Chen, Y. T., & Godfrey, M. (1996). Double mutant fibrillin-1 (FBN1) allele in a patient with neonatal Marfan syndrome. *J Med Genet*, 33(9), 760-763.
- Ward, C. J., Turley, H., Ong, A. C., Comley, M., Biddolph, S., Chetty, R., Ratcliffe, P. J., Gattner, K., & Harris, P. C. (1996). Polycystin, the polycystic kidney disease 1 protein, is expressed by epithelial cells in fetal, adult, and polycystic kidney. *Proc Natl Acad Sci U S A*, 93(4), 1524-1528.
- Waters, P. J., Parniak, M. A., Hewson, A. S., & Scriver, C. R. (1998). Alterations in protein aggregation and degradation due to mild and severe missense mutations (A104D, R157N) in the human phenylalanine hydroxylase gene (PAH). *Hum Mutat*, 12(5), 344-354.
- Watnick, T., & Germino, G. G. (1999). Molecular basis of autosomal dominant polycystic kidney disease. *Semin Nephrol*, 19(4), 327-343.
- Watnick, T., He, N., Wang, K., Liang, Y., Parfrey, P., Hefferton, D., St George-Hyslop, P., Germino, G., & Pei, Y. (2000). Mutations of PKD1 in ADPKD2 cysts suggest a pathogenic effect of trans-heterozygous mutations. *Nat Genet*, 25(2), 143-144.

- Watnick, T., Phakdeekitcharoen, B., Johnson, A., Gandolph, M., Wang, M., Briefel, G., Klinger, K. W., Kimberling, W., Gabow, P., & Germino, G. G. (1999). Mutation detection of PKD1 identifies a novel mutation common to three families with aneurysms and/or very-early-onset disease. *Am J Hum Genet*, 65(6), 1561-1571.
- Watnick, T. J., Gandolph, M. A., Weber, H., Neumann, H. P., & Germino, G. G. (1998). Gene conversion is a likely cause of mutation in PKD1. *Hum Mol Genet*, 7(8), 1239-1243.
- Watnick, T. J., Piontek, K. B., Cordal, T. M., Weber, H., Gandolph, M. A., Qian, F., Lens, X. M., Neumann, H. P., & Germino, G. G. (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(9), 1473-1481.
- Watnick, T. J., Torres, V. E., Gandolph, M. A., Qian, F., Onuchic, L. F., Klinger, K. W., Landes, G., & Germino, G. G. (1998). Somatic mutation in individual liver cysts supports a two-hit model of cystogenesis in autosomal dominant polycystic kidney disease. *Mol Cell*, 2(2), 247-251.
- Weber, B. L. (1998). Update on breast cancer susceptibility genes. *Recent Results Cancer Res*, 152, 49-59.
- Wen, L. (2001). Two-step cycle sequencing improves base ambiguities and signal dropouts in DNA sequencing reactions using energy-transfer-based fluorescent dye terminators. *Mol Biotechnol*, 17(2), 135-142.
- Weston, B. S., Jeffery, S., Jeffrey, I., Sharaf, S. F., Carter, N., Sagggar-Malik, A., & Price, R. G. (1997). Polycystin expression during embryonic development of human kidney in adult tissues and ADPKD tissue. *Histochem J*, 29(11-12), 847-856.
- White, M. B., Carvalho, M., Derse, D., O'Brien, S. J., & Dean, M. (1992). Detecting single base substitutions as heteroduplex polymorphisms. *Genomics*, 12(2), 301-306.
- Wijmenga, C., Padberg, G. W., Moerer, P., Wiegant, J., Liem, L., Brouwer, O. F., Milner, E. C., Weber, J. L., van Ommen, G. B., Sandkuyl, L. A., & et al. (1991). Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and in situ hybridization. *Genomics*, 9(4), 570-575.
- Wilson, P. D. (1997). Epithelial cell polarity and disease. *Am J Physiol*, 272(4 Pt 2), F434-442.
- Winship, P. R. (1989). An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. *Nucleic Acids Res*, 17(3), 1266.
- Woo, D. (1995). Apoptosis and loss of renal tissue in polycystic kidney diseases. *N Engl J Med*, 333(1), 18-25.

- Woo, D. D. L., Miao, S. Y. P., Pelayo, J. C., & Woolf, A. S. (1994). Taxol inhibits progression of congenital polycystic kidney disease. *Nature*, 368, 750-753.
- Wu, G., D'Agati, V., Cai, Y., Markowitz, G., Park, J. H., Reynolds, D. M., Maeda, Y., Le, T. C., Hou, H., Jr., Kucherlapati, R., Edelmann, W., & Somlo, S. (1998). Somatic inactivation of Pkd2 results in polycystic kidney disease. *Cell*, 93(2), 177-188.
- Wu, G., Mochizuki, T., Le, T. C., Cai, Y., Hayashi, T., Reynolds, D. M., & Somlo, S. (1997). Molecular cloning, cDNA sequence analysis, and chromosomal localization of mouse Pkd2. *Genomics*, 45(1), 220-223.
- Wu, G., & Somlo, S. (2000). Molecular genetics and mechanism of autosomal dominant polycystic kidney disease. *Mol Genet Metab*, 69(1), 1-15.
- Wunderle, V. M., Ramkisson, Y. D., Kwok, C., Korn, R. M., King, V. E., & Goodfellow, P. N. (1994). Breakpoint break for consortium studying adult polycystic kidney disease. *Cell*, 77(6), 785-786.
- Yamaguchi, T., Nagao, S., Takahashi, H., Ye, M., & Grantham, J. J. (1995). Cyst fluid from a murine model of polycystic kidney disease stimulates fluid secretion, cyclic adenosine monophosphate accumulation, and cell proliferation by Madin-Darby canine kidney cells in vitro. *Am J Kidney Dis*, 25(3), 471-477.
- Yen, P. M., Sugawara, A., Refetoff, S., & Chin, W. W. (1992). New insights on the mechanism(s) of the dominant negative effect of mutant thyroid hormone receptor in generalized resistance to thyroid hormone. *J Clin Invest*, 90(5), 1825-1831.
- Yersin, C., Bovet, P., Wauters, J. P., Schorderet, D. F., Pescia, G., & Paccaud, F. (1997). Frequency and impact of autosomal dominant polycystic kidney disease in the Seychelles (Indian Ocean). *Nephrol Dial Transplant*, 12(10), 2069-2074.
- Yoshida, H., Mitarai, T., Kawamura, T., Kitajima, T., Miyazaki, Y., Nagasawa, R., Kawaguchi, Y., Kubo, H., Ichikawa, I., & Sakai, O. (1995). Role of the deletion of polymorphism of the angiotensin converting enzyme gene in the progression and therapeutic responsiveness of IgA nephropathy. *J Clin Invest*, 96(5), 2162-2169.
- Youil, R., Kemper, B. W., & Cotton, R. G. (1995). Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII. *Proc Natl Acad Sci US A*, 92(1), 87-91.
- Yuan, B., Thomas, J. P., von Kodolitsch, Y., & Pyeritz, R. E. (1999). Comparison of heteroduplex analysis, direct sequencing, and enzyme mismatch cleavage for detecting mutations in a large gene, FBN1. *Hum Mutat*, 14(5), 440-446.

- Zarazaga, A., Garcia-De-Lorenzo, L., Garcia-Luna, P. P., Garcia-Peris, P., Lopez-Martinez, J., Lorenzo, V., Quecedo, L., & Del Llano, J. (2001). Nutritional support in chronic renal failure: systematic review. *Clin Nutr*, 20(4), 291-299.
- Zerres, K., Hansmann, M., Knopfle, G., & Stephan, M. (1985). Prenatal diagnosis of genetically determined early manifestation of autosomal dominant polycystic kidney disease? *Hum Genet*, 71(4), 368-369.
- Zerres, K., Rudnik-Schoneborn, S., & Deget, F. (1993). Childhood onset autosomal dominant polycystic kidney disease in sibs: clinical picture and recurrence risk. German Working Group on Paediatric Nephrology (Arbeitsgemeinschaft für Padiatrische Nephrologie. *J Med Genet*, 30(7), 583-588.
- Zhou, X. J., & Kukes, G. (1998). Pathogenesis of autosomal dominant polycystic kidney disease: role of apoptosis. *Diagn Mol Pathol*, 7(2), 65-68.
- Zielenski, J., Corey, M., Rozmahel, R., Markiewicz, D., Aznarez, I., Casals, T., Larriba, S., Mercier, B., Cutting, G. R., Krebsova, A., Macek, M., Jr., Langfelder-Schwind, E., Marshall, B. C., DeCelle-Germann, J., Claustres, M., Palacio, A., Bal, J., Nowakowska, A., Ferec, C., Estivill, X., Durie, P., & Tsui, L. C. (1999). Detection of a cystic fibrosis modifier locus for meconium ileus on human chromosome 19q13. *Nat Genet*, 22(2), 128-129.
- Zielenski, J., & Tsui, L. C. (1995). Cystic fibrosis: genotypic and phenotypic variations. *Annu Rev Genet*, 29, 777-807.

## **Appendices**



## **Appendices**

Appendix A

A1. PCR Optimisation with Differing buffers

Table A.1. Enzyme Buffers for Optimisation of Mg<sup>2+</sup> Concentrations

Series	Buffer Mix	Final [MgCl <sub>2</sub> ]mM
I	50 mM KCL	A = 0.75
	10 mM Tris-HCl pH 8.4	B = 1.50
	0.01 gelatin	C = 2.25
	0.1% Triton X-1000	D = 3.00
		E = 3.75
II	40 mM NaCl	F = 0.75
	10 mM KCl	G = 1.50
	0.01 gelatin	H = 2.25
	0.1% Triton X-1000	I = 3.00
		J = 3.75
III	10 mM (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	K = 0.75
	10 mM KCl	L = 1.50
	10 mM Tris-HCl pH 8.4	M = 2.25
	0.01 gelatin	N = 3.00
	0.1% Triton X-1000	O = 3.75

A2. QIAquick Spin Column Protocol

1. Initial PCR product was run on an agarose gel and the band excised and placed in a microcentrifuge tube (1.5ml).
2. 3 volumes of QX1 (QG) buffer were added to 1 volume of gel and it was melted at 50°C until dissolved.
3. Mixture was briefly vortexed, if pH indicator was yellow, 1 gel volume of Isopropanol was added and mixed.
4. Spin column was placed in 2ml collection tube and sample applied to column.
5. Sample was bound to the column by spinning at 13,000 rpm for 1 minute at room temp.

6. Flow through was discarded and 500µl of QX1 buffer applied to column, this was centrifuged again and flow through discarded.
7. 750µl of PE buffer was applied and column stood at rm. temp. for 2-5 minutes.
8. Column was centrifuged at 13,000 rpm for 1 minute, flow through discarded and centrifugation repeated.
9. Column was placed in a clean 1.5ml microcentrifuge tube.
10. DNA was eluted by the addition of 30µl of pure water to centre of membrane, this was allowed to stand for 1 minute and then centrifuged at 13,000rpm for 1 minute.

### A3. Centricon Filter Protocol

1. The remaining 90µl of ssDNA PCR product was applied to a Centricon YM30 column (cut off size approx 60 bp ssDNA)
2. 2.5ml of ddH<sub>2</sub>O was added to the column and centrifuged at 5,000g for 15 mins at room temp.
3. Flow through was discarded and column rinsed with 2.5ml of ddH<sub>2</sub>O, then centrifuged at 5,000g for 10 mins at room temp.
4. Step 3 was repeated twice and column was removed from lower discard tube
5. Column was inverted into a clean collecting tube and centrifuged at 3,500g for 5 minutes to collect concentrated ssDNA.

### A4. Cleaning of ABI Sequencing Reaction Prior to Electrophoresis

1. 2.5 volumes (25µl) of ice-cold 95% ethanol (EtOH) and 0.1 volumes (1µl) of 3M sodium acetate (pH 4.6) were added to the 10µl sequencing reaction.
2. The microcentrifuge tube was vortexed and left on ice or at -20°C for 30 minutes.

3. Tubes were then centrifuged at 13,000 rpm for 15-30 minutes at 4°C.
4. Supernatant was removed and care was taken not to disturb the pellet.
5. The pellet was rinsed in 250µl of 75% EtOH and then centrifuged at 13,000 rpm for 5 minutes at 4°C.
6. Supernatant was carefully removed and pellet dried to zero volume (either air-dried or dried in a rotary vacuum at room temperature).
7. Sequencing reactions were stored at -20°C prior to pellet being redissolved for electrophoresis.

#### A5. Plasmid Purification Post BAC/PAC clone identification

Purification of the culture was then performed using a modification of the Qiagen Plasmid Midi Kit protocol (personal communication Dr. Arseni Markoff, Muenster).

#### Maxi Prep from BAC/PAC DNA

1. Two days prior to isolation a single clone was picked from those grown on the LB Agar plate and used to inoculate 3-5ml of LB Media (containing chloramphenicol 12.5µg/ml) (LB/Cm) in a 15ml glass centrifuge tube. The culture was grown overnight at 37°C with vigorous shaking (~300rpm).
2. After 24 hours the starter culture was used to inoculate one litre of LB/Cm in a five litre flask and again shaken vigorously at 37°C overnight.
3. The following morning 500ml of the culture was transferred into 2 X 1L Beckmann centrifuge tubes and spun at 3,500 rpm for 10 mins at 4°C (JS horizontal rotor).

4. Supernatant was discarded and cells were resuspended in 10ml of P1 resuspension buffer (50mM Tris-Cl (pH 8.0), 10 mM EDTA, 100µg/ml RNase A, stored at 4°C) and vortexed.
5. 20ml of P2 lysis buffer (200mM NaOH, 1% SDS) was immediately added and the suspension was gently shaken then incubated at room temperature for 5 minutes.
6. 20ml of ice-cold P3 neutralisation buffer (3M potassium acetate pH 5.5) was added and tubes were mixed by inverting several times.
7. A Qiagen-tip 100 was equilibrated with 10ml of QBT equilibration buffer (750mM NaCl, 50 mM MOPS pH 7.0, 15% isopropanol, 0.15% Triton X-100).
8. The DNA precipitate was filtered through a funnel overlayed with cheese-cloth. The clear filtrate was then passed through the Qiagen tip in 2-3 shots by the application of vacuum.
9. The tip was then washed with 3 X 20ml of QC washing buffer (1.0M NaCl, 50 mM MOPS pH 7.0, 15% isopropanol) by applying vacuum.
10. The DNA was then eluted from the tip by the application of 5 X 1ml of pre-heated (65°C) QF elution buffer (1.25M NaCl, 50mM Tris, Tris-Cl pH 8.5, 15% isopropanol) into a 30ml Corex tube.
11. 0.7 X volumes (3.5ml) of isopropanol was then added and mixed well.
12. The tube was centrifuged at 10,000 rpm for 30 minutes at 4°C (JA 12 Sorvall rotor).
13. The supernatant was decanted and the pellet was washed with 10ml of 70% ethanol for 5 minutes at 10,000 rpm at 4°C.
14. The ethanol was then carefully decanted and the pellet air-dried. The pellet was resuspended in 500µl of TE buffer pH 8.0.

#### A6. ABI 373 DNA Sequencer (Perkin Elmer). Sample Preparation

1. A mixture of 5:1 deionised formamide (pH >7.0) to 25mM EDTA (pH 8.0), with 50mg/ml blue dextran was prepared.
2. 4µl of this mixture was added to each sequencing reaction.
3. Tubes were briefly vortexed, then centrifuged for 10 seconds.
4. Samples were denatured at 94°C for 3 minutes and placed immediately on ice.

#### Preparation of Sequencing Gel

These quantities make 60ml of 4% (19:1 acrylamide:bisacrylamide) gel with 8.3M urea in 1 X TBE, for a 48cm gel.

1. In a glass flask 30g urea, 22ml of ddH<sub>2</sub>O, 6ml of 40% acrylamide stock and approximately 1g of mixed bed resin were mixed.
2. The solution was stirred on a magnetic stirrer over a slight heat until the urea crystals began to dissolve and the flask was warm to touch. Stirring continued for 5-10 minutes without heating until all crystals were dissolved.
3. The acrylamide solution was filtered through a 0.2µm filter and the solution was degassed under vacuum for 5 minutes.
4. The gel mix was transferred to a 100ml cylinder and 6ml of 10 X TBE was added.
5. The mix was made up to a final volume of 60ml with ddH<sub>2</sub>O.
6. Clean plates were clamped together ready for gel pouring.
7. The gel solution was mixed in a beaker with 300µl of 10% ammonium persulphate and 30µl of TEMED. This was swirled gently to avoid air-bubbles being incorporated into the mix.

8. Plates were held at an angle, and gel was poured, when plates were full they were lowered to a horizontal position and an inverted sharks-tooth comb (48-well) was inserted to form the well trough.
9. The gel was allowed to polymerise at room temperature for 1-2 hours.

The comb was removed and the sequencing gel was assembled in the 373 DNA sequencer, the plate was scanned to ensure it was clean in the read-region with no fluorescent contamination. The sharks-tooth comb was re-inserted to form the wells. The upper and lower buffer chambers were then filled with 1 X TBE and pre-electrophoresed for five minutes at 40W to ensure that all connections were working prior to loading. Power was interrupted and the wells were rinsed with buffer using a syringe, 2µl of each denatured sample was loaded and the gel was electrophoresed at 40W constant power for 12-15 hours. Sequencing data were then saved to disk for analysis by requestor.

#### A7. Sequitherm Excel II Cycle Sequencing

##### Cycle Sequencing Protocol

##### Premix

Primer 5pmol	1 µl
[α <sup>33</sup> P ]-dATP	0.5µl (10µCi)
5X SequiTherm EXCEL Buffer	5µl
DNA Template (200 fmoles)	5µl
ddH <sub>2</sub> O (to 16µl)	4.5µl
5U of SequiTherm EXCEL II enzyme	1.0µl

For each template four 600µl microcentrifuge tubes were labelled A, G, T and C and 2µl of the appropriate SequiTherm EXCEL II Termination mix was placed into each tube on ice. 4µl of the premix solution was then pipetted into each tube and mixed thoroughly. The tubes were denatured at 95°C for five minutes and the reactions were cycled 30 times at 95°C for 15 seconds, 65°C for 1 minute and 70°C for 1 minute. After 30 cycles 5µl of stop buffer (95% vol/vol formamide, 20 mM EDTA, 1% xylene cyanol FF, 1% bromophenol blue) was added to each reaction tube.

## **Appendix B**

Journal articles resulting directly from this thesis.

1. Bogdanova & McCluskey *et al.* (2000). Screening the 3' region of the polycystic kidney disease 1 (PKD1) gene in 41 Bulgarian and Australian kindreds reveals a prevalence of protein truncation mutations. *Hum Mutat.* 16(2): 166-174.
2. Bogdanova *et al.* (2001). Homologues to the first gene for polycystic kidney disease are pseudogenes. *Genomics.* 74(3): 333-341.
3. McCluskey *et al.* (2002). Mutation detection in the duplicated region of the polycystic kidney disease 1 (PKD1) gene in PKD1-linked Australian families. *Hum Mutat.* 19(3): 240-250.



## **Appendix B1,**

Bogdanova & McCluskey *et al.* (2000). Screening the 3' region of the polycystic kidney disease 1 (PKD1) gene in 41 Bulgarian and Australian kindreds reveals a prevalence of protein truncation mutations. *Hum Mutat.* 16(2): 166-174.

## RESEARCH ARTICLE

# Screening the 3' Region of the Polycystic Kidney Disease 1 (PKD1) Gene in 41 Bulgarian and Australian Kindreds Reveals a Prevalence of Protein Truncating Mutations

Nadja Bogdanova,<sup>1</sup> Marie McCluskey,<sup>1,2</sup> Karin Sikmann,<sup>1</sup> Arseni Markoff,<sup>1</sup> Vassil Todorov,<sup>3</sup> Dimitar Dimitrakov,<sup>4</sup> Tina Schiavello,<sup>2</sup> Mark Thomas,<sup>5</sup> Luba Kalaydjieva,<sup>2</sup> Bernd Dworniczak,<sup>1</sup> and Jürgen Horst<sup>1\*</sup>

<sup>1</sup>Institut für Humangenetik, Westfälische Wilhelms-Universität Münster, Münster, Germany

<sup>2</sup>Centre for Human Genetics, Edith Cowan University, Joondalup, Australia

<sup>3</sup>Clinic of Nephrology and Haemodialysis, Medical University, Pleven, Bulgaria

<sup>4</sup>Clinic of Nephrology and Haemodialysis, Medical University, Plovdiv, Bulgaria

<sup>5</sup>Department of Nephrology, Royal Perth Hospital, Perth, Australia

Communicated by Michel Gossens

Screening for disease-causing mutations in the unique region of the polycystic kidney disease 1 (PKD1) gene was performed in 41 unrelated individuals with autosomal dominant polycystic kidney disease. Exons 34-41 and 43-46 were assayed using PCR amplification and SSCP analysis followed by direct sequencing of amplicons presenting variant SSCP patterns. We have identified seven disease-causing mutations of which five are novel [c.10634-10656del; c.11587delG; IVS37-10C>A; c.11669-11674del; c.13069-13070ins39] and two have been reported previously [Q4010X; Q4041X]. Defects in this part of the gene thus account for 17% of our group of patients. Five of the seven sequence alterations detected are protein-truncating which is in agreement with mutation screening data for this part of the gene by other groups. The two other mutations are in-frame deletions or insertions which could destroy important functional properties of polycystin 1. These findings suggest that the first step toward cyst formation in PKD1 patients is the loss of one functional copy of polycystin 1, which indirectly supports the "two-hit" model of cystogenesis where a second somatic mutation inactivating the normal allele is necessary to occur for development of the disease condition. *Hum Mutat* 16:166-174, 2000. © 2000 Wiley-Liss, Inc.

**KEY WORDS:** PKD1; polycystic kidney disease; autosomal dominant; polycystin 1; genotype-phenotype correlation

## DATABASES

PKD1 – OMIM:601313, 173900 (PKD); GDB:120293; GenBank:L39891; HGMD:PKD1

## INTRODUCTION

Autosomal dominant polycystic kidney disease (ADPKD) is one of the most common Mendelian inherited disorders (MIM# 173900) with an incidence of approximately one in 1,000 [Dalgard, 1957]. The main clinical manifestation of the disease is the development of fluid-filled cysts in both kidneys which, in most affected individuals, results in kidney failure. Eight to 10% of all patients requiring haemodialysis or renal transplantation have ADPKD [Gabow, 1993]. Additional symptoms may include hypertension, urinary tract infections,

Received 23 February 2000; accepted revised manuscript 26 April 2000.

\*Correspondence to: Jürgen Horst, Inst. für Humangenetik, Westfälische Wilhelms-Universität Münster, 48149 Münster, Germany.

Contract grant sponsor: Deutsche Forschungsgemeinschaft; Contract grant number: Ho 572/18-2; Contract grant sponsor: Interdisziplinäres Zentrum für Klinische Forschung, Münster; Contract grant number: IZKF2 B4; Contract grant sponsor: Edith Cowan University; Contract grant number: 99-6354; Contract grant sponsor: Medical Research Foundation, Royal Perth Hospital; Contract grant number: 41-96.

Nadja Bogdanova and Marie McCluskey contributed equally to this work.

haematuria, proteinuria, extra-renal cysts, cerebral aneurysms, herniae, cardiac valve defects, renal calculi, gastrointestinal diverticuli [Roscoe et al., 1993], and spinal meningeal diverticula [Schievink and Torres, 1997]. The clinical course of ADPKD varies in terms of presenting symptoms and the rate of progression towards kidney failure. In some patients end-stage renal failure (ESRF) occurs in adolescence while in others ESRF does not occur at all [Choukroun et al., 1995; Parfrey et al., 1990]. Among those who develop ESRF the mean age at onset is approximately 50 years [Milutinovic et al., 1984]. The wide range of clinical manifestations and the variation in severity are most probably a compound product of the genetic heterogeneity, locus as well as allelic, within ADPKD itself and of the effects of modifying genes and environment.

At least three different genes are currently implicated in autosomal dominant polycystic kidney disease: PKD1 (MIM# 601313), located on chromosome 16p13.3 [The European Polycystic Kidney Disease Consortium, 1994], PKD2 (MIM# 173910) on chromosome 4q21-23 [Mochizuki et al., 1996], and at least one more gene which is yet to be identified [Bogdanova et al., 1995; de Almeida et al., 1995; Daoust et al., 1995].

The participation of additional environmental and genetic factors in the course of the disease is supported by the demonstration of a "two-hit" mechanism in renal- and hepatic-cyst formation. Germline and somatic inactivating mutations of the two characterized genes for ADPKD, PKD1, and PKD2 have been identified in cyst-lining epithelial cells [Qian et al., 1996; Watnick et al., 1998; Koptides et al., 1999]. To study the genotype-phenotype correlation and address the functional importance of predicted protein domains further efforts have been made to screen PKD1 and PKD2 for disease-causing alterations of the nucleotide and amino acid sequence.

In affected individuals of European descent, mutations in PKD1 are the most common cause for the disease and account for up to 85% of cases [Peters and Sandkuijl, 1992]. The PKD1 gene consists of 46 exons spanning ~52 kb of genomic DNA [The International Polycystic Kidney Disease Consortium, 1995; The American PKD1 Consortium, 1995; Hughes et al., 1995]. A large part of the gene is duplicated in at least three homologous genes, HG-A, HG-B, and HG-C, which are located on chromosome 16p13.1 and share approximately 95–97% homology with the PKD1 gene [Hughes et al., 1995]. Only 3.5 kb of the PKD1 transcript (ap-

proximately 14 kb in length), located at the 3' end of the gene (exons 34 to 46), is unique to PKD1. The PKD1 gene product, polycystin 1, consists of 4,302 amino acids and is predicted to contain a large N-terminal extracellular domain of ~2,500 residues, several transmembrane domains, and a short cytoplasmic C-terminal region [The International Polycystic Kidney Disease Consortium, 1995; Hughes et al., 1995; Sandford et al., 1997]. It has been demonstrated recently that polycystin 1 and the gene product of PKD2, polycystin 2, associate *in vivo* in a heterodimeric interaction [Qian et al., 1997] involving the carboxy-termini of the two proteins. In polycystin 1, the proposed region of interaction is positioned in exon 46, the final exon of the gene (nucleotide positions 12854–12955 of the cDNA, acc. L33243). To date, a total of 40 mutations in the 3' end of PKD1 have been reported, mostly detected in patients from western European countries.

Here we present the results of a collaborative study of mutations in the unique region of PKD1 in 41 Bulgarian and Australian ADPKD individuals, which indicate predominant appearance of protein-truncating sequence alterations in this part of the gene.

## MATERIALS AND METHODS

### Patients

Twenty-six of the affected subjects are from Bulgaria and represent a mixed southeastern European population. The Australian group consists of 15 affected individuals predominantly of British descent with one family of Spanish origin. In all patients, the diagnosis of ADPKD was based on standard ultrasound criteria [Bear et al., 1992]. Linkage studies have been conducted in 39 of the families. Two affected individuals, one from Australia (A72.1) and one Bulgarian (B25.1), were included in the study without prior linkage analysis, based on the assumption that PKD1 accounts for most cases of ADPKD.

### DNA Isolation and PCR Amplification

DNA was extracted from peripheral whole blood using standard procedures [Miller et al., 1988]. Polymerase chain reaction (PCR) was performed on 50 ng of genomic DNA in a final volume of 25  $\mu$ l using 1.5 units of Taq polymerase (Gibco BRL Life Technologies) and 1.5 mM  $MgCl_2$ . The optimised conditions for the PCR amplification of individual exons are shown in Table 1. Prior to SSCP, the PCR products were electrophoresed in 1% agarose gels and visualized with ethidium

TABLE 1. Amplification Conditions for SSCP Analysis of the Unique 3' Region of PKD1

Amplification primers	Amplicon size (bp)	Annealing and cycling conditions	Additives	Nucleotide position*
<b>Exon 34</b>	240	30 cycles at 65°C	BSA (0.5%) gelatine (2%)	44276-295 44496-515
Forward 5' gtgagctgggtgagaggag 3'				
Reverse 5' agggctgctggcctgagtc 3'				
<b>Exon 35</b>	220	2-step PCR of 30 cycles at 68°C	Non	47330-349 47530-549
Forward 5' ctgcaactgctctcaggagg 3'				
Reverse 5' ctgacatctccctgtgagct 3'				
<b>Exon 36</b>	286	30 cycles at 63°C	Non	47541-560 47807-826
Forward 5' ctgtgagctgctctcagag 3'				
Reverse 5' ctacaggctcctcctcagg 3'				
<b>Exon 37</b>	299	2-step PCR of 35 cycles at 68°C	Non	47801-820 48080-099
Forward 5' ggtaggctcaggctcct 3'				
Reverse 5' cactcagctctctgtctcc 3'				
<b>Exon 38</b>	272	30 cycles at 64°C	Non	48439-458 48691-710
Forward 5' caaagctctgctctcagct 3'				
Reverse 5' tagtctagctcagctcagct 3'				
<b>Exon 39</b>	259	2-step PCR of 35 cycles at 68°C	Non	48929-948 49167-187
Forward 5' gtctctgggtgctcctcct 3'				
Reverse 5' gtagctctagctcagctct 3'				
<b>Exon 40</b>	264	2-step PCR of 35 cycles at 68°C	Non	49330-349 49573-593
Forward 5' gaggctgctcagctcagct 3'				
Reverse 5' ctacagctcagctcagct 3'				
<b>Exon 41</b>	355	2-step PCR of 35 cycles at 68°C	Non	49571-590 49906-925
Forward 5' cgtctcagctcagctcagct 3'				
Reverse 5' ctctcagctcagctcagct 3'				
<b>Exon 43</b>	383	40 cycles at 60°C	DMSO (2%)	50381-403 50741-763
Forward 5' cagcgtcctcagctcagct 3'				
Reverse 5' gctcagctcagctcagct 3'				
<b>Exon 44</b>	292	2-step PCR of 35 cycles at 68°C	Non	50746-765 51018-037
Forward 5' caccctcagctcagctcagct 3'				
Reverse 5' gtgagctcctctcagctcagct 3'				
<b>Exon 45</b>	396	30 cycles at 68°C	Non	50946-965 51322-341
Forward 5' cgtctagctcagctcagct 3'				
Reverse 5' agtgggctcagctcagct 3'				
<b>Exon 46 (fragment a)</b>	318	2-step PCR of 35 cycles at 68°C	Non	51361-380 51659-678
Forward 5' gtgagctcagctcagctcagct 3'				
Reverse 5' cgtctcagctcagctcagct 3'				
<b>Exon 46 (fragment b)</b>	336	35 cycles at 66°C	DMSO (2%)	51633-652 51950-968
Forward 5' gacgagctcagctcagctcagct 3'				
Reverse 5' caggcagctcagctcagctcagct 3'				

\*According to Genbank accession number L39891.

bromide staining in order to confirm amplification and rule out large deletions or insertions.

### SSCP Analysis

Single-strand conformation polymorphism (SSCP) analysis was performed using 2 µl of PCR product and an equal volume of formamide loading buffer (98% formamide, 0.025% xylene cyanol, 0.025% bromophenol blue). The mix was heat-denatured and kept on ice prior to loading. We found that the optimal gel composition for electrophoresis was a 12% polyacrylamide gel with a cross-linking ratio (%C) equal to 2.5 [Savov et al., 1992]. The gels (160 × 200 × 0.75 mm) were run at 25°C for varying lengths of time at different constant voltages using a vertical gel apparatus, D Gene™ Sys-

tem (Biorad), in 1 × TAE buffer without glycerol and in 1 × TBE with 5% glycerol [Kukita et al., 1997]. The single strand separation was visualized by a silver-staining procedure [Budowle et al., 1991]. The gels were dried and stored for documentation.

### DNA Sequencing and Restriction Analysis

All samples where an aberrant banding pattern had been detected by SSCP analysis were subjected to manual sequencing using internal labelling with [ $\alpha$ -<sup>32</sup>P]-dATP (Hartmann Analytics). Products for the sequencing reactions were obtained by PCR on genomic DNA with the amplification primers listed in Table 1. Cycle sequencing in forward and reverse directions using the same or internal primers was performed with the SequiTherm™ Cycle Sequenc-

ing Kit according to the manufacturer's protocol (Epicentre Technologies). The products of the sequencing reactions were electrophoresed in 6% denaturing polyacrylamide gels (C%=5; 6M urea; 350 × 400 × 0.4mm), at 68W constant power for 2–3 hr. Gels were dried under vacuum prior to autoradiographic exposure on Biomax™ films (Kodak).

Restriction digestion of gel purified PCR products was performed in a total volume of 20 µl with 10 units of the appropriate enzyme (Gibco BRL; New England Biolabs), in conditions specified by the manufacturer. Reaction products were separated on a 2.5% agarose gel and visualized by ethidium bromide staining.

All sequence changes were subject to family segregation analysis using SSCP or, where possible, restriction fragment length polymorphism analysis.

## RESULTS

Using the techniques outlined above, we have identified disease-causing mutations in the unique 3' region of the PKD1 gene in seven of the 41 ADPKD patients screened (summarized in Tables 2 and 3). Five novel and two previously reported mutations were detected. The results are presented according to the type of molecular change.

## Protein Truncating Mutations

Two nonsense mutations were detected in this study, both of which have been reported previously.

### Q4041X

This is a C to T substitution at nucleotide c.12332 (Genbank acc. L33243) which changes a Glutamine to a Stop codon and is predicted to truncate polycystin 1 by 262 amino acids. It was first identified by Turco et al. [1995] in an Italian patient and subsequently reported in one Spanish and two British families [Torra et al., 1998; Daniells et al., 1998].

In our study, Q4041X was found in individual A3.1, the index patient from a large Australian family of British origin, and was also detected in all six affected family members participating in the study. Hepatic cysts have been detected in 3/6 affected individuals. Apart from that, the clinical course of ADPKD, including rate of progression towards renal failure, appears variable, as shown in Table 2.

### Q4010X

This is a C to T substitution at nucleotide c.12239 which changes a Glutamine to a Stop codon and is predicted to truncate polycystin 1 by 293 amino acids. It has been reported previously in

TABLE 2. Clinical Data of Patients From 5 Bulgarian and 2 Australian PKD1 Families With Disease-Causing Mutations in the Unique Part of the PKD1 Gene Identified in the Present Study

Mutation (exon/intron)	Family & patient ID	Sex	Current age (years)	CRF <sup>a</sup> (years)	ESRF <sup>b</sup> (years)	Hypertension	Liver cysts	Intracranial aneurysms
c.10634-10656del (Exon 34)	B22.1	M	63	42	49	Yes	No	No
	B22.2	F	61	54	57	Yes	No	No
	B22.3	F	26	No	No	No	No	No
	B22.4	F	38	No	No	No	No	No
	B22.5	F	34	No	No	No	No	No
IVS 37-10C>A (Intron 37)	B57.1	F	64	54	55	Yes	Yes	No
	B57.2	F	64	55	No	No	Yes	No
	B57.3	M	42	39	No	Yes	No	No
c.11587delG (Exon 40)	B77.1	F	55	45	No	Yes	Yes	No
	B77.2	F	55	49	No	Yes	Yes	No
	B77.3	M	31	No	No	Yes	Yes	No
	B77.4	M	29	No	No	Yes	No	No
	B77.5	M	28	No	No	No	No	No
c.11669-11674del (Exon 41) Q4010X (Exon 44) Q4041X (Exon 44)	B77.6	M	25	No	No	No	No	No
	B25.1	M	45	44	No	Yes	No	No
	A72.1	M	53	45	49	Yes	No	No
	A3.1	M	60	?	42	Yes	?	?
	A3.11	F	66	50	57	Yes	Yes	Yes
	A3.13	M	45	No	No	Yes	Yes	No
	A3.18	M	41	No	No	No	No	No
	A3.19	F	39	No	No	No	No	No
	A3.33	M	51	No	No	No	No	No
	B41.1	F	35	No	No	Yes	Yes	No
c.51815-5181ins39 (Exon 46)	B41.2	M	39	No	No	No	No	No

<sup>a</sup>CRF, chronic renal failure.

<sup>b</sup>ESRF, end stage renal failure.

TABLE 3. Disease-Causing Mutations in the Unique Part of the PKD1 Gene Identified in the Present Study

Exon/ intron	Site of nucleotide change <sup>a</sup>	Amino acid change	Restriction site change	Reference
Exon 34	c.10634-10656del	P34575X3617	-AvaII	Present study
IVS37	IVS37-10C>A	Splice site	+BsrI	Present study
Exon 40	c.11587delG	R37935X3824		Present study
Exon 41	c.11669-11674del	QE3820-3821del	-AluI	Present study
Exon 44	c.12239C>T	Q4010X	-HhaI	Turco et al., 1995
	c.12332C>T	Q4041X	+MaeI	Daniells et al., 1998
Exon 46	c.13069-13070ins39	A286-4287ins13as		Present study

<sup>a</sup>According to Genbank accession number L39243.

an ADPKD patient of British descent [Daniells et al., 1998] and in a Korean patient [Kim et al., 2000].

In our study, the mutation was found in Australian patient A72.1. No other family members were available for this study. The family has lived in Australia for the past three generations and originates from Britain. Patient A72.1 was diagnosed at 15 years of age due to positive family history. He developed hypertension at age 27 and reached ESRF at 49 years (Table 2). His other symptoms include umbilical hernia and gout.

Three of the novel mutations identified in this study are predicted to result in a frameshift and subsequent introduction of a preliminary stop codon. Two of these changes are deletions and one is a possible splice mutation likely to have a frameshift effect.

#### c.11587delG.

A single base pair deletion in exon 40 was found in Bulgarian patient B77.1. The mutation is predicted to result in termination of translation at codon 3824 and to truncate polycystin 1 by 478 amino acids.

The presence of this mutation was tested by the use of SSCP analysis in eight other family members, of whom six are known to have ADPKD. Analysis of the other family members showed that the c.11587delG mutation segregated with the disease. The clinical features of ADPKD in this family are shown in Table 2.

#### c.10634-10656del.

This novel 23 bp deletion in exon 34 results in a stop codon at position 3617 thus truncating the protein by 685 amino acids. The mutation was found in Bulgarian patient B22.1 and in all five affected members of the family. Their phenotypic features are shown in Table 2.

#### Splice-Site Mutation (IVS37-10C>A)

A transversion of C to A at position g.48491 (Genbank accession number L39891) in intron 37

of PKD1 was detected in Bulgarian patient B57.1. This change could create a new acceptor splice site eight nucleotides upstream of the end of IVS37 as shown below.

Wild Type	48481 agggccccc cgtccac [cag]
	AGC CTC CTG . . .
Mutant	48481 agggccccc [cag] TCC ACC
	AGA GCC TCC TGG . . .

The new acceptor site produces a frameshift with a translation stop at codon 3684 thus resulting in truncation of polycystin 1 by 618 amino acids.

The nucleotide substitution creates a Bsr I restriction site and restriction fragment length polymorphism analysis was used to screen additional family members and 50 unrelated normal subjects for the presence of the mutation. It was found that the mutation co-segregates with the disease in this family and it is not present in the normal controls (data not shown). Thus, we presume that this could indeed be a disease-causing mutation, although no RNA from this patient was available to verify that. The disease manifestation in the affected family members is summarized in Table 2.

#### In-frame Deletions and Insertions

##### c.11669-11674del.

This 6 bp deletion in exon 41, resulting in the loss of glutamine and glutamic acid at codons 3820-3821, was observed in Bulgarian index patient B25.1. Overview of the disease phenotype in the affected patient is shown in Table 2.

There were no other family members available so no segregation analysis has been performed. However, 50 unrelated normal individuals were tested for the presence of this deletion, using genomic amplification and subsequent separation of the PCR products on a 3% agarose gel. No carrier of the mutation was detected in the control group (data not shown), thus suggesting that this could be a disease-causing mutation.

In order to relate this mutation to an eventual

function distortion of polycystin 1, the protein segment encoded by exon 41 was subjected to a secondary structure prediction analysis, PHDsec [Rost and Sander, 1994], with and without the deletion change. Figure 1 shows the comparison between both variants; the deletion of glutamine and glutamic acid at positions 3820 and 3821 modifies the topology of this portion of the protein, which is a predicted part of a loop exposed on the external cell surface [Hughes et al., 1995].

**c.13069-13070lms39.**

An in-frame insertion in exon 46 of 39 bp was detected in patient B41.1. The insertion begins at nucleotide 13069, between codons 4286 (Pro) and 4287 (Ser) as detailed below.

(c.13067)CCC ^acg cag gga cca ctt cgg gcc  
aag aac aag gtc cac ccc ^ AGC(c.13072).

These 39 nucleotides encode the following additional amino acids:

(4286) Pro ^ Thr Gln Gly Pro Leu Arg Ala Lys  
Asn Lys Val His Pro ^ Ser(4287).

Ten of the 13 inserted amino acids represent a repeated segment further downstream in the sequence (positions 4290–4299). This insertion creates a third additional loop at the very C-terminus of the protein, as shown by using the PHDsec algorithm (Fig. 2).

The insertion was detected in the two members of the family who are known to be affected and in two young individuals (born in 1987 and 1986) who are predicted to be carriers of a PKD1

polycykin I, pos. 5044 - 5345

## C.4L. endnotes

[illegible]

841, c1169-1174

amino acids	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
sec. structure	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
prediction	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
subset	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

## polycystin 1, par. 4349 - 4381

**2.46. 1994. 1995. 1996. 1997. 1998. 1999. 2000. 2001. 2002. 2003. 2004. 2005. 2006. 2007. 2008. 2009. 2010. 2011. 2012. 2013. 2014. 2015. 2016. 2017. 2018. 2019. 2020. 2021. 2022. 2023. 2024. 2025. 2026. 2027. 2028. 2029. 2030. 2031. 2032. 2033. 2034. 2035. 2036. 2037. 2038. 2039. 2040. 2041. 2042. 2043. 2044. 2045. 2046. 2047. 2048. 2049. 2050. 2051. 2052. 2053. 2054. 2055. 2056. 2057. 2058. 2059. 2060. 2061. 2062. 2063. 2064. 2065. 2066. 2067. 2068. 2069. 2070. 2071. 2072. 2073. 2074. 2075. 2076. 2077. 2078. 2079. 2080. 2081. 2082. 2083. 2084. 2085. 2086. 2087. 2088. 2089. 2090. 2091. 2092. 2093. 2094. 2095. 2096. 2097. 2098. 2099. 2100. 2101. 2102. 2103. 2104. 2105. 2106. 2107. 2108. 2109. 2110. 2111. 2112. 2113. 2114. 2115. 2116. 2117. 2118. 2119. 2120. 2121. 2122. 2123. 2124. 2125. 2126. 2127. 2128. 2129. 2130. 2131. 2132. 2133. 2134. 2135. 2136. 2137. 2138. 2139. 2140. 2141. 2142. 2143. 2144. 2145. 2146. 2147. 2148. 2149. 2150. 2151. 2152. 2153. 2154. 2155. 2156. 2157. 2158. 2159. 2160. 2161. 2162. 2163. 2164. 2165. 2166. 2167. 2168. 2169. 2170. 2171. 2172. 2173. 2174. 2175. 2176. 2177. 2178. 2179. 2180. 2181. 2182. 2183. 2184. 2185. 2186. 2187. 2188. 2189. 2190. 2191. 2192. 2193. 2194. 2195. 2196. 2197. 2198. 2199. 2200. 2201. 2202. 2203. 2204. 2205. 2206. 2207. 2208. 2209. 2210. 2211. 2212. 2213. 2214. 2215. 2216. 2217. 2218. 2219. 2220. 2221. 2222. 2223. 2224. 2225. 2226. 2227. 2228. 2229. 2230. 2231. 2232. 2233. 2234. 2235. 2236. 2237. 2238. 2239. 2240. 2241. 2242. 2243. 2244. 2245. 2246. 2247. 2248. 2249. 2250. 2251. 2252. 2253. 2254. 2255. 2256. 2257. 2258. 2259. 2260. 2261. 2262. 2263. 2264. 2265. 2266. 2267. 2268. 2269. 2270. 2271. 2272. 2273. 2274. 2275. 2276. 2277. 2278. 2279. 2280. 2281. 2282. 2283. 2284. 2285. 2286. 2287. 2288. 2289. 2290. 2291. 2292. 2293. 2294. 2295. 2296. 2297. 2298. 2299. 2300. 2301. 2302. 2303. 2304. 2305. 2306. 2307. 2308. 2309. 2310. 2311. 2312. 2313. 2314. 2315. 2316. 2317. 2318. 2319. 2320. 2321. 2322. 2323. 2324. 2325. 2326. 2327. 2328. 2329. 2330. 2331. 2332. 2333. 2334. 2335. 2336. 2337. 2338. 2339. 2340. 2341. 2342. 2343. 2344. 2345. 2346. 2347. 2348. 2349. 2350. 2351. 2352. 2353. 2354. 2355. 2356. 2357. 2358. 2359. 2360. 2361. 2362. 2363. 2364. 2365. 2366. 2367. 2368. 2369. 2370. 2371. 2372. 2373. 2374. 2375. 2376. 2377. 2378. 2379. 2380. 2381. 2382. 2383. 2384. 2385. 2386. 2387. 2388. 2389. 2390. 2391. 2392. 2393. 2394. 2395. 2396. 2397. 2398. 2399. 2400. 2401. 2402. 2403. 2404. 2405. 2406. 2407. 2408. 2409. 2410. 2411. 2412. 2413. 2414. 2415. 2416. 2417. 2418. 2419. 2420. 2421. 2422. 2423. 2424. 2425. 2426. 2427. 2428. 2429. 2430. 2431. 2432. 2433. 2434. 2435. 2436. 2437. 2438. 2439. 2440. 2441. 2442. 2443. 2444. 2445. 2446. 2447. 2448. 2449. 2450. 2451. 2452. 2453. 2454. 2455. 2456. 2457. 2458. 2459. 2460. 2461. 2462. 2463. 2464. 2465. 2466. 2467. 2468. 2469. 2470. 2471. 2472. 2473. 2474. 2475. 2476. 2477. 2478. 2479. 2480. 2481. 2482. 2483. 2484. 2485. 2486. 2487. 2488. 2489. 2490. 2491. 2492. 2493. 2494. 2495. 2496. 2497. 2498. 2499. 2500. 2501. 2502. 2503. 2504. 2505. 2506. 2507. 2508. 2509. 2510. 2511. 2512. 2513. 2514. 2515. 2516. 2517. 2518. 2519. 2520. 2521. 2522. 2523. 2524. 2525. 2526. 2527. 2528. 2529. 2530. 2531. 2532. 2533. 2534. 2535. 2536. 2537. 2538. 2539. 2540. 2541. 2542. 2543. 2544. 2545. 2546. 2547. 2548. 2549. 2550. 2551. 2552. 2553. 2554. 2555. 2556. 2557. 2558. 2559. 2560. 2561. 2562. 2563. 2564. 2565. 2566. 2567. 2568. 2569. 2570. 2571. 2572. 2573. 2574. 2575. 2576. 2577. 2578. 2579. 2580. 2581. 2582. 2583. 2584. 2585. 2586. 2587. 2588. 2589. 2590. 2591. 2592. 2593. 2594. 2595. 2596. 2597. 2598. 2599. 2600. 2601. 2602. 2603. 2604. 2605. 2606. 2607. 2608. 2609. 2610. 2611. 2612. 2613. 2614. 2615. 2616. 2617. 2618. 2619. 2620. 2621. 2622. 2623. 2624. 2625. 2626. 2627. 2628. 2629. 2630. 2631. 2632. 2633. 2634. 2635. 2636. 2637. 2638. 2639. 2640. 2641. 2642. 2643. 2644. 2645. 2646. 2647. 2648. 2649. 2650. 2651. 2652. 2653. 2654. 2655. 2656. 2657. 2658. 2659. 2660. 2661. 2662. 2663. 2664. 2665. 2666. 2667. 2668. 2669. 2670. 2671. 2672. 2673. 2674.**

```

amino acids      |ALPGLIARAKRIVYLATQPSLITFLAKDQIVFST|
sec. structure   |                                      |
prediction
subset           |11...          ...11...11111111...11111111|

```

WAS, c.13069-13070b.33

```
amino acids |ALPGLARASQVILATVLTORPLRAKIVITFETPLRAKIVITFET|
sec. structure |XXXXXXXXXXXXX|
prediction
subset |LL..H.HN...LL...LTLTLL.....LTLTTLTLL.....LTLTTLTLL
```

**FIGURE 2.** Alignment of the C-terminal 35 amino acids of polycystin 1, ex. 46 and the C-terminal 48 amino acids of ex. 46, c.13069-13070ins39 and secondary structure prediction using the PHDsec algorithm. The 13 inserted residues are underlined. Repeated amino acids are shown in italic case. The symbols for secondary structure are the same as in Figure 1.

mutation on the basis of the linkage analysis with polymorphic markers flanking the PKD1 gene. An outline of the clinical findings in this family is provided in Table 2.

## DISCUSSION

Mutation detection in the PKD1 gene is complicated by the high GC content of the gene (around 62%) and, most of all, by the interference of the homologous genes. Predictably, most mutation detection studies have therefore focused on the unique 3' region of PKD1. Of the 82 mutations reported to date [Human Genome Mutation Database], 45 (55% of all mutations detected) are in the 3' unique region which, however, accounts for only 19% of the coding sequence. When these data are viewed as a true proportion of all individuals tested (approx 970 to date) [500 (British, Italian and Spanish) Daniells et al., 1998; 175 (Spanish) Badenas et al., 1999; 15 (Cypriots) Neophytou et al., 1996; 146 (French) Perrichot et al., 1999; 90 (British) Afzal et al., 1999; 41 (Australian and Bulgarian) present paper], only 5–6% of individuals have mutations in this region. In the present study, we have detected mutations in 16% of the ADPKD individuals investigated (12.5% Australian and 18.5% Bulgarian) over the same region. Both sets of data suggest that there is no clustering of mutations in the unique part of the PKD1 gene. Expressed as a percentage of patients investigated, the number of disease-causing mutations in the 3' end of the gene is roughly proportional to its length, therefore the majority of PKD1 mutations should remain located in the duplicated region.

**FIGURE 1.** Alignment of polycystin 1, exon 41, and exon 41, c.11669-11674del amino acid sequences and secondary structure prediction using the PHDsec algorithm. The deleted amino acids are underlined in the wild type sequence. E means extended sheet, H means helix and L means loop. The prediction subset in both figures is meaningful for all residues with an expected average correlation greater than 0.69.

Given the sequence characteristics of PKD1 and its complex genomic structure, designing a mutation screening strategy is not easy. Our experimental protocol, using SSCP as the screening technique, has been successful in identifying seven disease-causing mutations in 41 individuals tested. Thus, the proportion of identified mutations in our study is at least as high as that in other studies of this part of the gene. Five out of the seven sequence alterations detected would lead to a premature termination of translation. They were all identified through SSCP analysis, which is clearly not a method biased towards this type of mutation.

The exons analyzed here in the 3' region did not include exon 42 of the gene. Despite repeated attempts at amplification, using two different sets of oligonucleotide primers, we failed to obtain a specific product. No mutations have been reported to date in this exon, suggesting that other groups may also experience similar difficulties with specific amplification. This could be due to the presence of a non-disease-causing polymorphism in IVS 42. It includes a 34 bp repeat region and produces an intron size variation of ~325 bp to ~130 bp [Peral et al., 1996a]. This variation may explain the problems we encountered in the amplification of this part of the gene.

Out of the total of 50 mutations identified in the unique part of PKD1 so far (including the present study), only five, namely Q4041X [Turco et al., 1995; Daniells et al., 1998; Torra et al., 1998; Badenas et al., 1999], Q4010X [Daniells et al., 1998; Kim et al., 2000], R4020X [Rossetti et al., 1996; Neophytou et al., 1998], R4227X [Peral et al., 1996a; Peral et al., 1997], and R4275W [Badenas et al., 1999] have been reported to occur in more than one family. Two of these, namely Q4041X and Q4010X, were also detected in the present study. Comparison of polymorphic haplotypes of the Q4041X-bearing chromosomes in the Italian, British, and Spanish families with this mutation [Torra et al., 1998], revealed no identity by descent. A slipped-mispairing model explaining the recurrence of this mutation was subsequently proposed by Daniells et al. [1998]. We cannot rule out identity by descent for either Q4041X or Q4010X. The two mutations have been identified previously in British patients [Daniells et al., 1998] and in our study were detected in Australian individuals of British ancestry. Whatever the mechanism of origin, the most common PKD1 mutation known to date, Q4041X, accounts for less than 1% (five out of 877) of disease alleles. The emerging pattern is that of a large

multitude of private mutations confined to individual families, thus complicating further the study of molecular defects and of genotype-phenotype correlations in ADPKD.

Regardless of the variety of mutations detected, preliminary conclusions are already possible on the predominating type of defects. Of the reported mutations in the 3' end of PKD1 [Human Genome Mutation Database, Cardiff], approximately 80% are protein truncating. This is in contrast with the findings in the 5' part of the gene, where 37% (13/35) of the defects so far reported are missense mutations resulting in amino acid substitutions [Human Genome Mutation Database, Cardiff]. The same trend was observed in our study, where five of the seven disease-causing mutations detected in the unique part of the PKD1 gene would lead to a truncated protein. Moreover, missense mutations were identified in our studies but failed to segregate with the disease and were thus considered to be polymorphic variations (manuscript in preparation). It should be noted, however, that our study was performed on multiplex families where the disease penetrance is high and linkage analysis indicating PKD1 as the gene involved in ADPKD was possible. The spectrum of mutations could be different in the whole disease population which contains many isolated families with too few cases to make linkage analysis possible and where the disease penetrance may be low. Nevertheless, the protein-truncating mutations identified in our study cause C-terminal disruption of polycystin 1 prior to the last 113 amino acids, which would render the protein incapable of interaction with polycystin 2 [Tsiokas et al., 1997; Qian et al., 1997]. These findings suggest that the first step toward cyst formation in PKD1 patients is the loss of one functional copy of polycystin 1, which indirectly supports the "two-hit" model of cystogenesis where a second somatic mutation inactivating the normal allele is necessary to occur for development of the disease condition.

The novel in-frame deletion and insertion we found in this study seem to be pathological mutations, not only as judged by their non-occurrence in healthy subjects. The secondary structure prediction data for the relevant segments of the protein do show a potential functional interference caused by these changes. c.11669-11674del is perhaps distorting a possible interaction of polycystin 1 with an extracellular ligand by changing the topology of the second large extracellular loop, according to the protein model proposed by Hughes et al. [1995]. c.13069-13070ins39 is clearly chang-



ing the spatial organization of the very C-terminus of the protein by introducing an additional loop. This mutation causes an early onset of ADPKD in the assayed family. The last 115 amino acids of the cytoplasmic C-terminus of polycystin 1 seem to be essential for modulation of Wnt signaling during renal development [Kim et al., 1999] but the proposed potential ligand(s) for cytoplasmic-signalling cascades have not yet been identified. A conformation change as gross as the one induced by this insertion could definitely influence the interacting properties of this portion of the protein. The inserted repeated stretch of 10 amino acids, on the other hand, could be a consequence of a more general slipped-mispairing mutation generating mechanism.

An assessment of the phenotypic effects of mutations in PKD1 is rendered difficult by the limited clinical data provided in mutation reports. We have been able to find information on the age at onset of ESRF for 21 families with 19 different mutations in the unique region of PKD1, including this study [Peral et al., 1996a, 1996b, 1997; Turco et al., 1995; Neophytou et al., 1996; Torra et al., 1998; The European Polycystic Kidney Disease Consortium, 1994]. The total number of affected individuals in these families was 39. The mean age at onset of ESRF in patients with mutations in the transmembrane domain ( $N=28$ ) was 53.8 years and for patients with mutations in the cytoplasmic region ( $N=11$ ) it was 52.7 years. No significant differences could be detected when mean age at onset of ESRF was compared between groups of patients with different types of mutation. It was 53.2 years for patients with truncating mutations ( $N=30$ ) and 54.5 years for non-truncating mutations ( $N=9$ ). A meaningful assessment of the clinical effect of mutations in different exons is not possible because of the very small numbers of patients with mutations in the same exon. As already suggested by the comparison between individuals with mutations in the transmembrane domain and those with mutations in the cytoplasmic part of polycystin 1, most reach ESRF about age 55 regardless of the specific exon affected.

Since the causes of clinical variation in ADPKD are of great interest and since allelic variation may be one such cause, it is of interest to compare the phenotype characteristics of individuals with the same molecular defect. We have therefore examined the information on ADPKD patients with Q4041X. Hepatic cysts appear to be commonly associated with this mutation and have been re-

ported to occur in eight of 12 individuals on whom clinical information is available (including this study). The rate of deterioration of kidney function is variable, with 2/5 individuals developing ESRF before age 45 and 3/5 after 55 years. The presence of hypertension has been reported in seven of 10 individuals over the age of 30. This variation points further to the complexity of the disorder and supports again the "two hit" model suggesting that germline mutations in the PKD genes are only one of many determinants in the development and clinical spectrum of ADPKD.

### ACKNOWLEDGMENTS

Marie McCluskey and Tina Schiavello are recipients of scholarships from the Faculty of Communications, Health and Science at Edith Cowan University, Western Australia.

### REFERENCES

- Abzal AR, Hand M, Ternes-Pereira E, Sagger-Malik A, Taylor R, Jeffrey S. 1999. Novel mutations in the 3' region of the polycystic kidney disease 1 (PKD1) gene. *Hum Genet* 105: 648-653.
- The American Polycystic Kidney Disease 1 Consortium. 1995. Analysis of the genomic sequence for the autosomal dominant polycystic kidney disease gene predicts the presence of a leucine-rich repeat. *Hum Mol Genet* 4:575-582.
- Badenas C, Torra R, San Millan JL, Lucero L, Milla M, Estivill X, Darnell A. 1999. Mutational analysis within the 3' region of the PKD1 gene. *Kidney Int* 55:1225-1233.
- Bear JC, Parfrey PS, Morgan JM, Martin CJ, Cramer BC. 1992. Autosomal dominant polycystic kidney disease: new information for genetic counselling. *Am J Med Genet* 43: 548-553.
- Bogdanova N, Dworniczak B, Dragova D, Todorov V, Dimitrakov D, Kalinov K, Hallmayer J, Horst J, Kalaydjieva L. 1995. Genetic heterogeneity of polycystic kidney disease in Bulgaria. *Hum Genet* 95:645-650.
- Budowie B, Chakraborty R, Giusti AM, Eisenberg AJ, Allen RC. 1991. Analysis of the VNTR Locus DIS80 by the PCR followed by High-Resolution PAGE. *Am J Hum Genet* 48: 137-144.
- Choukroun G, Itakura Y, Men NK, Christophe JL, Allouze G, Jungers P, Grunfeld JR. 1995. The rate of progression of renal failure in ADPKD. *Contrib Nephrol* 115:28-32.
- Dalgard OZ. 1957. Bilateral polycystic kidney disease of the kidneys: a follow-up of two hundred and eighty-four patients and their families. *Acta Med Scand* 328:1-251.
- Daniels 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.
- Daoust MC, Reynolds DM, Bicher DG, Soule S. 1995. Evidence for a third genetic locus for autosomal dominant polycystic kidney disease. *Genomics* 25:733-736.
- de Almeida S, de Almeida E, Peters D, Pinto JR, Tavora I, Lavinha J, Breuning M, Prata MM. 1995. Autosomal dominant polycystic kidney disease: evidence for the existence of a third locus in a Portuguese family. *Hum Genet* 96:83-88.
- The European Polycystic Kidney Disease Consortium. 1994. The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. *Cell* 77:881-894.

- Gabow PA. 1993. Autosomal dominant polycystic kidney disease. *N Engl J Med* 329:332-342.
- Hughes J, Ward CJ, Peral B, Aspinwall R, Clark K, San Millan JL, Gamble V, Harris PC. 1995. The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. *Nature Genet* 10:151-160.
- The Human Gene Mutation Database, Cardiff. <http://www.uwcm.ac.uk/uwcm/tsg/search/120293.html>
- The International Polycystic Kidney Disease Consortium. 1995. Polycystic kidney disease: the complete structure of the PKD1 gene and its protein. *Cell* 81:289-298.
- Kim E, Arnould T, Sellin LK, Benzing T, Fan MJ, Gruning W, Sokol SY, Drummond I, Walz G. 1999. The polycystic kidney disease 1 gene product modulates Wnt signalling. *J Biol Chem* 274:4947-4953.
- Kim UK, Jin DK, Ahn C, Shin JH, Lee KB, Kim SH, Chae JJ, Hwang DY, Lee JG, Namkoong Y, Lee CC. 2000. Novel mutations of the PKD1 gene in Korean patients with autosomal dominant polycystic kidney disease. *Mutat Res* 432:39-45.
- Koptides M, Hadjimichael C, Koupepidou P, Pierides A, Constantinou Deltas C. 1999. Germinal and somatic mutations in the PKD2 gene of renal cysts in autosomal dominant polycystic kidney disease. *Hum Mol Genet* 8:509-513.
- Kukita Y, Tahira T, Sommer SS, Hayashi K. 1997. SSCP analysis of long DNA fragments in low pH gel. *Hum Mutat* 10:400-407.
- Miller SA, Dykes DD, Polesky HF. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucl Acids Res* 16:12-15.
- Milutinovic J, Fialkow PJ, Agodoa LY, Phillips LA, Rudd TG, Bryant JL. 1984. Autosomal dominant polycystic kidney disease: symptoms and clinical findings. *Q J Med* 53:511-522.
- Mochizuki T, Guanqing W, Hayashi T, Xenophontos SL, Veldhuisen B, Saris JJ, Reynolds DM, Cai Y, Gabow PA, Pierides A, Kimberling WJ, Breuning MH, Deltas CC, Peters DJM, Somlo S. 1996. PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science* 272:1339-1342.
- Neophytou P, Constantinides R, Lazarou A, Pierides A, Deltas CC. 1996. Detection of a novel nonsense mutation and an intragenic polymorphism in the PKD1 gene of a Cypriot family with autosomal dominant polycystic kidney disease. *Hum Genet* 98:437-442.
- Neophytou P, Constantinides R, Girginoudis P, Papapavliou P, Koptides M, Ioannou P, Eleftheriou A, Papadopolou E, Papadopoulou D, Loucopoulos D, Demetriou K, Pierides A, Deltas CC. 1998. Identification of novel and recurrent mutations in the polycystic kidney disease 1 gene (PKD1) by single strand conformation analysis. *Balkan J Med Genet* 1:149-159.
- Parfrey PS, Bear JC, Morgan J, Cramer BC, McManamon PJ, Gault MH, Churchill DN, Singh M, Hewitt R, Somlo S, Reeders ST. 1990. The diagnosis and prognosis of autosomal dominant polycystic kidney disease. *N Engl J Med* 323:1085-1090.
- Peral B, San Millan JL, Ong AC, Gamble V, Ward CJ, Strong C, Harris PC. 1996a. Screening the 3' region of the polycystic kidney disease 1 (PKD1) gene reveals six novel mutations. *Am J Hum Genet* 58:86-96.
- Peral B, Ong AC, San Millan JL, Gamble V, Rees L, Harris PC. 1996b. A stable, nonsense mutation associated with a case of infantile onset polycystic kidney disease 1 (PKD1). *Hum Mol Genet* 5:539-542.
- Peral B, Gamble V, Strong C, Ong AC, Sloane-Stanley J, Zerres K, Winearls CG, Harris PC. 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.
- Perrichot RA, Mercier B, Simon PM, Whebe B, Cledes J, Ferec C. 1999. DGGE screening of PKD1 gene reveals novel mutations in a large cohort of 146 unrelated patients. *Hum Genet* 105:231-239.
- Peters DJ, Sandkuijl LA. 1992. Genetic heterogeneity of polycystic kidney disease in Europe. *Contrib Nephrol* 97:128-139.
- 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.
- Qian F, Germino FJ, Cai Y, Zhang X, Somlo S, Germino GG. 1997. PKD1 interacts with PKD2 through a probable coiled-coil domain. *Nature Genet* 16:179-183.
- Roscoe JM, Brissenden JE, Williams EA, Chery AL, Silverman M. 1993. Autosomal dominant polycystic kidney disease in Toronto. *Kidney Int* 44:1101-1108.
- Rossetti S, Bresin E, Restagno G, Carbonara A, Corra S, De Prisco O, Pignatti PF, Turco AE. 1996. Autosomal dominant polycystic kidney disease (ADPKD) in an Italian family carrying a novel nonsense mutation and two missense changes in exons 44 and 45 of the PKD1 Gene. *Am J Med Genet* 65:155-159.
- Rost B, Sander C. 1994. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* 20:55-72.
- Sandford R, Sgotto B, Aparicio S, Brenner S, Vaudin M, Wilson RK, Chissoc S, Pepin K, Bateman A, Chothia C, Hughes J, Harris P. 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.
- Savov A, Angelicheva D, Jordanova A, Eigel A, Kalsaydjieva L. 1992. High percentage acrylamide gels improve resolution in SSCP analysis. *Nucl Acid Res* 20:6741-6742.
- Schievink WI, Torres VE. 1997. Spinal meningeal diverticula in autosomal dominant polycystic kidney disease. *Lancet* 349:1223-1224.
- Torra R, Badenas C, Peral B, Darnell A, Serra E, Gamble V, Turco AE, Harris PC, Estivill X. 1998. Recurrence of the PKD1 nonsense mutation Q401X in Spanish, Italian, and British families. *Hum Mutat Suppl.* 1:S117-S120.
- Tsiokas L, Kim E, Arnould T, Sukhatme VP, Walz G. 1997. Homo- and heterodimeric interactions between the gene products of PKD1 and PKD2. *Proc Natl Acad Sci USA* 94:6965-6970.
- Turco AE, Rossetti S, Bresin E, Corra S, Gammato L, Maschio G, Pignatti PF. 1995. A novel nonsense mutation in the PKD1 gene (C3187T) is associated with autosomal dominant polycystic kidney disease (ADPKD) in a large three-generation Italian family. *Hum Mol Genet* 4:1331-1335.
- Watnick TJ, Torres VE, Gandolph MA, Qian F, Onuchic LF, Klingler KW, Landes G, Germino GG. 1998. Somatic mutation in individual liver cysts supports a two-hit model of cystogenesis in autosomal dominant polycystic kidney disease. *Mol Cell* 2:247-251.

## **Appendix B2**

Bogdanova *et al.* (2001). Homologues to the first gene for polycystic kidney disease are pseudogenes. *Genomics*. 74(3): 333-341.

# Homologues to the First Gene for Autosomal Dominant Polycystic Kidney Disease Are Pseudogenes

Nadia Bogdanova,<sup>\*,1</sup> Arseni Markoff,<sup>†,1</sup> Volker Gerke,<sup>†</sup> Marie McCluskey,<sup>‡</sup> Jürgen Horst,<sup>\*,2</sup> and Bernd Dworniczak<sup>\*</sup>

<sup>\*</sup>Institut für Humangenetik and <sup>†</sup>Institut für Medizinische Biochemie, ZMBE, Westfälische Wilhelms-Universität Münster, D-48149 Münster, Germany; and <sup>‡</sup>Centre for Human Genetics, Edith Cowan University, Jondaloo, Australia

Received January 29, 2001; accepted April 6, 2001

PKD1 is the first gene identified to be causative for the condition of autosomal dominant polycystic kidney disease. There are several genes homologous to PKD1 that are located proximal to the master gene on the same chromosome. Two of these genes have been recently covered in a large sequencing work on chromosome 16, and their structure has been broadly analyzed. However, the major question whether homologous genes (HG) code for functionally active polypeptides has not been resolved so far. The current study identifies and partially characterizes four more homologues of PKD1, different from the previously published sequence, two of which were found by screening of a BAC library and the other two contained in available databases. Analysis of HG transcripts shows that they are not translated in the model cell line T98G. Taken together, these findings suggest that homologues to PKD1 form a family of pseudogenes. © 2001 Academic Press

## INTRODUCTION

PKD1 is the first gene found to be impaired in autosomal dominant polycystic kidney disease (ADPKD). This condition is one of the most common monogenic inherited disorders with an incidence of approximately 1 in 1000 (Dalgaard, 1957). Genetic linkage studies in affected families show that at least three different genes are implicated in ADPKD: PKD1, located on chromosome 16p13.3 (The European Polycystic Kidney Disease Consortium, 1994); PKD2 on chromosome 4q21–q23 (Mochizuki *et al.*, 1996); and at least one more gene, which is yet to be identified (Bogdanova *et al.*, 1995; De Almeida *et al.*, 1995; Daoust *et al.*, 1995).

Sequence data from this article have been deposited with the GenBank Data Library under Accession Nos. AF320593 and AF320594.

<sup>1</sup> Both first authors contributed equally to this work.

<sup>2</sup> To whom correspondence should be addressed at the Institut für Humangenetik, WWU Münster, Vesaliusweg 12-14, D-48149 Münster, Germany. Telephone: +49 251 835-5401. Fax: +49 251 835-5431. E-mail: [horstj@uni-muenster.de](mailto:horstj@uni-muenster.de).

Germ-line and somatic inactivating mutations of the two characterized genes for ADPKD, PKD1 and PKD2, have been identified in cyst-lining epithelial cells (Qian *et al.*, 1996; Watnick *et al.*, 1998; Koptides *et al.*, 1999), providing evidence for a “two hit” mechanism causing cyst formation.

In affected individuals of European descent, mutations in PKD1 are the most common cause for the disease and account for up to 85.5% of cases (Peters and Sandkuijl, 1992). The PKD1 gene extends over ~52 kb of genomic DNA and contains 46 exons encoded by a 14-kb transcript (The International Polycystic Kidney Disease Consortium, 1995; The American PKD1 Consortium, 1995; Hughes *et al.*, 1995). The PKD1 gene product, polycystin 1, consists of 4302 amino acids and is predicted to contain a large N-terminal extracellular domain of ~2500 residues, several transmembrane domains, and a short cytoplasmic C-terminal region (The International Polycystic Kidney Disease Consortium, 1995; Hughes *et al.*, 1995; Sandford *et al.*, 1997). It has been demonstrated that polycystin 1 and the gene product of PKD2, polycystin 2, associate *in vivo* in a heterodimeric interaction (Qian *et al.*, 1997; Tsiokas *et al.*, 1997) involving the carboxy-termini of the two proteins. This interaction of both polycystins results in the formation of a new calcium-permeable nonselective cation channel recently described (Hanaoka *et al.*, 2000). The complexly organized extracellular N-terminus of polycystin 1 is thought to be involved in cell–cell or cell–matrix interactions.

A large part of the PKD1 gene is duplicated in an unknown number of homologous genes (HG), which are located on chromosome 16p13.1 and share approximately 95–97% homology with the PKD1 gene (Hughes *et al.*, 1995). Only 3.5 kb of the PKD1 transcript, located at the 3' end of the gene, is unique to PKD1. The exact number of the HG as well as their structure is yet unknown. More important, although published data indicate that these genes produce large transcripts (The European Polycystic Kidney Disease Consortium, 1994), it is still unknown whether they would code for functional polypeptides. An answer to this question is

TABLE 1

Primers and Annealing Temperatures for Second Round Amplification of Exons 6–14 of the PKD1 Gene

Exon	Primer sequence	Genomic position	Annealing temperature (°C)	Product size (bp)
6	6F 5' gtgagtgctgctgca 3'	21491–21508	58	396
	6R 5' ctcttctctctgagact 3'	21868–21886		
7	7F 5' ggctctgagcctcagttt 3'	22104–22121	65	391
	7R 5' taaccacagccagcgtct 3'	22476–22493		
8	8F 5' gctcttctctctgctgct 3'	22575–22592	58	306
	8R 5' ccacttctctctgctgct 3'	22863–22880		
9	9F 5' gctcttctctctgctgct 3'	23090–23108	60	271
	9R 5' gtgaaagctcagagaggcca 3'	23341–23360		
10	10F 5' gctgagctgctgctgctgct 3'	23608–23627	61	340
	10R 5' gctgagagagatgcagggga 3'	23929–23947		
11–11a	11aF 5' gggctccagggcctgacctg 3'	24300–24320	66	331
	11aR 5' cgcagctcagggctgctgct 3'	24610–24630		
11–11b	11bF 5' acggaacagctcagctgct 3'	24581–24600	66	320
	11bR 5' aaggctggccaacagggcaggg 3'	24880–24900		
11–11c	11cF 5' gctgctcagcgcctgctgct 3'	24849–24869	66	321
	11cR 5' agcaactctctctgctgct 3'	25150–25170		
12	12F 5' ccaggagcgagagggcta 3'	25923–25940	61	273
	12R 5' ggtagaggtgaaggtgga 3'	26178–26196		
13	13F 5' tcacctgcaactgggtcac 3'	26279–26299	68	261
	13R 5' gacaagagctggtgcccacc 3'	26519–26539		
14	14F 5' ctgtccggtcactcactgc 3'	26769–26789	68	250
	14R 5' cagctgactgggagctggg 3'	26998–27018		

of substantial interest regarding the gene function and the gene evolution in particular, because it has been shown that there are no Pkd1 homologues in mouse (Olsson *et al.*, 1996).

Up to now mutation screens of PKD1 have been heavily obstructed by the presence of these highly homologous copies of the gene extending from exon 1 to the first 87 bp of exon 33. Although different selective strategies for analysis of the duplicated part of the PKD1 gene have been employed successfully by several groups, methods used, such as the long-range amplification-based approaches, present a serious challenge for routine mutation detection. Therefore, the accumulation of precise sequence information on the HG is desperately required to design PKD1-specific reagents that would serve more conventional scanning techniques.

By screening a bacterial artificial chromosome (BAC) DNA library, we have identified clones that contain exclusively PKD1-homologous genes. Further analysis of these clones allowed us to characterize the near 5' region of homologues and to design primers for specific amplification of exons 6–14 of the PKD1 gene. Differences identified by comparing the sequences of these HG with the PKD1 master gene allowed us also to amplify specifically segments of PKD1 and of the PKD1-like genes from total and translated cDNA pools. Results obtained clearly indicated that PKD1 homologues are transcribed but not translated in the glioblastoma cell line T98G. In addition, detailed analysis of the HG coding sequences demonstrates the presence of premature stop codons and thus, if translated, they would generate only short, most probably non-functional polypeptides. Our conclusion from this com-

parison is that these genes are pseudogenes, duplicated in the course of molecular evolution.

## MATERIALS AND METHODS

**BAC and PAC library screening and characterization.** Primer pairs from the PKD1 unique and duplicated regions were used for identification of PKD1 or PKD1-homologous genes containing BAC and PAC clones. The BAC (Human BAC DNA Pools, Release II; Research Genetics, Inc., Huntsville, AL) and PAC (Genome Systems) libraries were screened through PCR according to the manufacturers' instructions. Positive BAC and PAC clones were grown and DNA was isolated as previously described (Pennekamp *et al.*, 1998). Clones were sequenced for genomic regions covering exon 3 through to intron 14 in direct or amplicon sequencing reactions using PKD1 exon-specific primers in both directions. Sequencing reactions were performed with the Applied Biosystems (ABI) BigDye Terminator Cycle Sequencing Ready Reaction Kit according to the manufacturer's guidelines and analyzed on an ABI 310 automated sequencer. Obtained sequences were compared with each other and with the PKD1 genomic sequence.

**PCR amplification on genomic DNA.** PKD1-specific templates were generated for exons 6–11 and 12–14 by amplification from genomic DNAs with the following primers: for exons 6–11, F1 primer 5' GGCCGCTACAGCATCGTGGCCC 3', position 21447–21470 (PKD1 genomic sequence, GenBank Accession No. L39891), and R1 primer 5' CAGTGAGGGGAGGCACCTACACTG 3', position 25622–25644, and for exons 12–14, F2 primer 5' CAGTGTAGGTGCCTCCCTCACTGCT 3', position 25621–25644, and R2 primer 5' CCAGCGGTCAGGAGGTACACTC 3', position 27410–27430. PCR was 35 cycles of denaturation at 94°C for 1 min, annealing at 62°C for 1 min, and extension at 72°C for 4 min. Reactions were performed in a total volume of 25 µl and the reaction mix contained the following: 60 mM Tris-Cl, pH 9.5, 20 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1 mM DTT, 0.005% NP-40, 1.5 mM MgCl<sub>2</sub>, 0.5 M betaine, 5% DMSO, 20 pM each primer, 600 µM dNTP, 100–200 ng DNA, and 1.25 U Taq polymerase produced and purified as previously described (Pluthero, 1993). PCR products of 4.2 and 1.8 kb accordingly were obtained and after dilution of 10<sup>-4</sup> they were used in second rounds of amplifications with primers specific for each exon to be analyzed (Table 1). Reaction

conditions were 30 cycles of denaturation at 94°C for 1 min, annealing (see Table 1) for 1 min, and extension at 72°C for 1 min.

**Reverse transcription and PCR (RT-PCR).** Total RNA was isolated from confluent glioblastoma T98G cell cultures by using the Trizol (Gibco BRL Life Technologies) reagent. Translationally active RNA was prepared from a polysomal fraction (membrane-bound polysomes) as previously described (Mechler, 1987). Briefly, the microsomal fraction was obtained through sucrose gradient centrifugation of a T98G cytoplasmic extract and mRNA was separated and purified from the bound polysomes. Total RNA and mRNA (0.3–0.5 µg) were subjected to reverse transcription using the SuperScript II reverse transcriptase and reaction conditions specified by the manufacturer (Gibco BRL Life Technologies). Reactions were primed with 100 pM random hexamer and 10 pM 3' oligonucleotide primer common for PKD1 and homologous genes, 5' GCCATCCCCGAAAG-GTCCAGTCGAA 3', position 4187–4230 (PKD1 cDNA sequence, GenBank Accession No. L33243), located in exon 15 of the PKD1 gene. First round of PCR amplification was performed on 1-µl aliquots of the RT reaction, with oligonucleotides RT-F1, 5' GTGTGC-CCGTCCTCGGTGCAGA 3', position 1294–1316 (exon 5), and RT-R1, 5' CCGAAGTCCACGTGTAAAGAA 3', position 3681–3702 (exon 15). Cycling conditions were denaturation at 94°C for 45 s, annealing at 55°C for 1 min, and extension at 72°C for 4 min in 40 cycles. Aliquots (1 µl) of the first PCR were used in a second amplification round with "nested" oligonucleotide primers specific for PKD1 and for HG. PKD1 reactions were primed with oligonucleotides RT-F2-PKD1, 5' CTGAGGCCCAACCTGGACT 3', position 2560–2780 (exon 11), and RT-R2-PKD1, 5' CCGGCACTGTG-GAGACCTGCAGA 3', position 3274–3296 (exon 13). HG reactions were performed with primers RT-R2-HG, 5' CCTGAGGCCAAC-CCTGGGCT 3', and 5' CTGCACTGTAGAGACCGGCAGG 3' (differences with the PKD1 sequence are underlined). PCR was 40 cycles of denaturation at 94°C for 45 s, annealing at 68°C for 1 min, and extension at 72°C for 1 min. Reaction products were resolved on EtBr-stained 1% agarose gels.

**Sequencing of PCR products.** The sequence of RT-PCR amplicons from PKD1 and homologous genes was determined with the amplification primers listed above.

First- and second-round amplicons from human genomic DNA were sequenced and compared with the PKD1 genomic sequence. All sequencing reactions were performed by using amplification primers and the Sequitherm Excel II DNA Sequencing Kit as described by the manufacturer (Epicentre Technologies), and labeling was with [ $\alpha$ -<sup>32</sup>P]dATP (Hartmann Analytics). The products of the sequencing reactions were electrophoresed in 6% denaturing polyacrylamide gels (C% = 5; 6 M urea; 350 × 400 × 0.4 mm), at 70 W constant power for 2–3 h. Gels were dried under vacuum prior to autoradiographic exposure on Biomax films (Kodak).

**Computer analysis.** Additional BAC clones containing other PKD1 homologous genes were found by using BLAST search on the GenBank and EMBL databases. The 7.7-kb region sequenced from our HG-containing BAC clones was used for the query. Sequence alignments and degree of homology estimates were accomplished by using the GENESTREAM global alignment tool. Sequenced regions were analyzed for repeats and low-complexity DNA at the Repeat-Masker Web Server.

## RESULTS

### Structural Analysis of Homologous Genes

Two positive BAC clones (204\_o\_15 and 267\_d\_9) containing PKD1-homologous genes and a third clone, containing a part of PKD1 as shown later by sequencing, amplified with primers from the repeated region and did not amplify with the primers located in the PKD1 unique region. The sequence comparison with the PKD1 genomic sequence and with each other for intron regions in introns 3, 4, 7, 8, 11, 12, 13, and 14

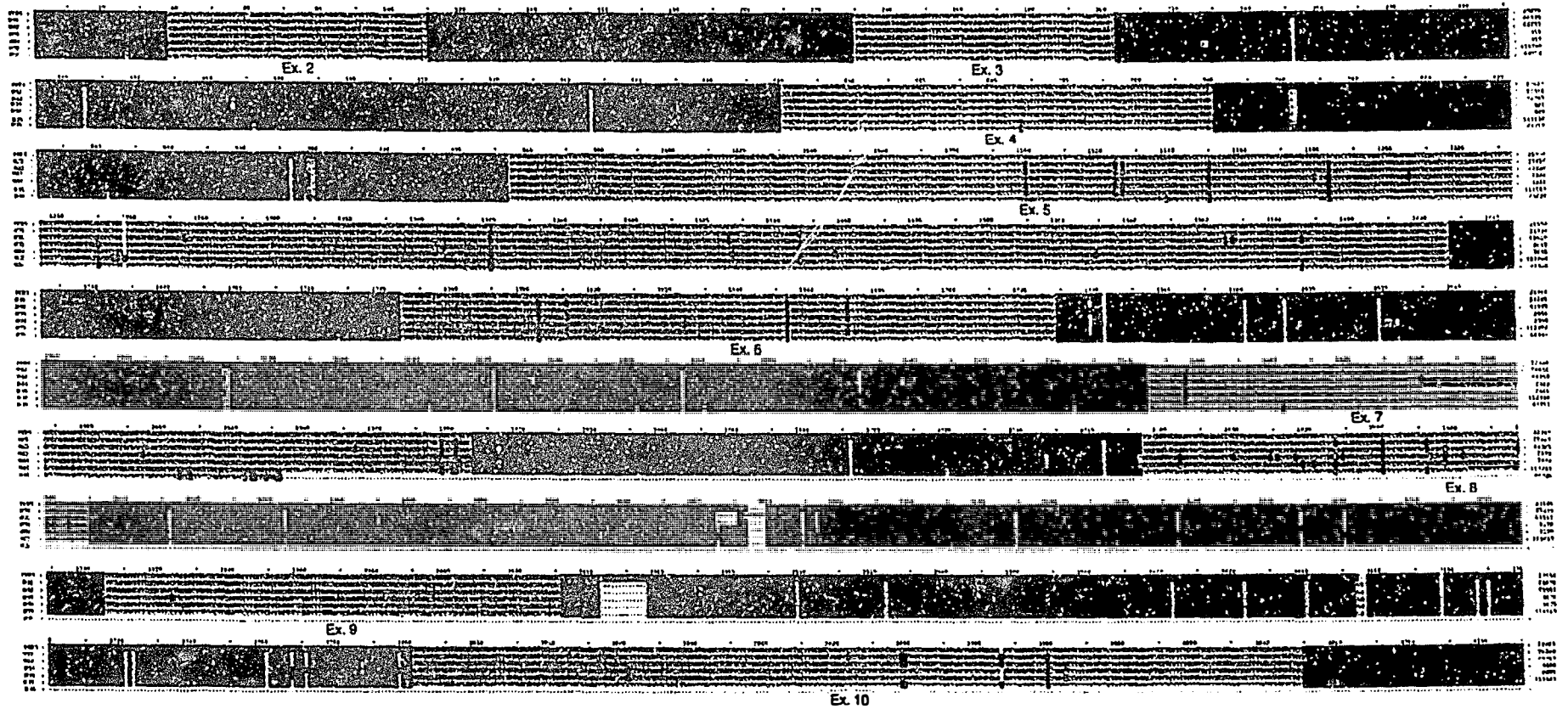
showed that both BAC clones contain different PKD1-homologous gene sequences. These clones were completely sequenced for the region corresponding to position 19866 through position 27595 (exon 2 to IVS 14) of the PKD1 gene. The sequences obtained (7.7 kb) were queried against the GenBank and EMBL databases and four more BAC clones showing high degrees of homology to the queried segment were identified. Further analysis showed that two of those clones (GenBank Accession Nos. AC009065 and AC012171) contain the PKD1 gene or a part thereof (overall sequence identity 99.6 and 99.8%) and two other clones contain PKD1-homologous genes.

To answer the question whether the PKD1 homologous genes we identified are the same genes found in the BAC clone A-13F4 described by Loftus *et al.* (1999), we performed a sequence alignment of those genes with the genes contained in our BAC clones, as well as with the homologous gene sequences identified through our BLAST search, on regions corresponding to position 19866–27595 of the PKD1 genomic sequence (Fig. 1). Overall sequence identity for each of the homologous gene-containing BACs with the PKD1 genomic sequence was 95.4–96.7% for the sequenced regions. There were, however, several bigger deletions/insertions (5–21 bp) identified in introns 9, 11, 12, and 13.

The result of this sequence comparison of every possible homologous gene couple, including the PKD1 sequence, is summarized in Table 2. The comparison clearly indicates that we have identified in our BAC clones two more different homologous genes, which we term PKD1P3 and PKD1P4 (PKD1 pseudogenes 3 and 4), and another two homologous genes contained in publicly available BAC clone sequences (GenBank Accession Nos. AC010488 and AC040158). The latter genes we term PKD1P5 and PKD1P6 accordingly. PKD1P3 (GenBank Accession No. AF320593) is contained in BAC clone 204\_o\_15, and clone 267\_d\_9 harbors the PKD1P4 gene (GenBank Accession No. AF320594). Preliminary data on those clones show that PKD1P3 and PKD1P4 begin to share homology with the PKD1 genomic sequence in the first third of intron 1, in the PKD1P5 sequence this is position 4150, and in PKD1P6, position 6566. Both genomic positions land at potential splice acceptor sites. PKD1P3 and PKD1P4 seem to end their PKD1-homologous region with exon 31, PKD1 homology in PKD1P5 ends at position 34235, in the beginning of the polypyrimidine tract found in intron 21. The PKD1P6 homology with PKD1 extends over only 2530 bp after position 19866, before it ends with position 22393 in exon 7 at a potential splice acceptor site. That is why we exclude this gene from quantitative comparisons in the above defined 7.7-kb region.

The sequence homology levels with PKD1 show that PKD1P2 (previously HG2, Loftus *et al.*, 1999) is next to identical with PKD1, whereas PKD1P1 (previously HG1) and PKD1P5 are farthest distant and PKD1P3 and PKD1P4 are in the middle, but closer to PKD1P2.

# Alignment of PKD1 and its pseudogenes



**FIG. 1.** Alignment of PKD1 and available homologous gene sequences, exon 2–IVS 14. Exons are shaded in gray and introns are boxed in black. Sequence discrepancies within exonic regions are marked through black boxes, sequence variations in intronic regions are highlighted in white.



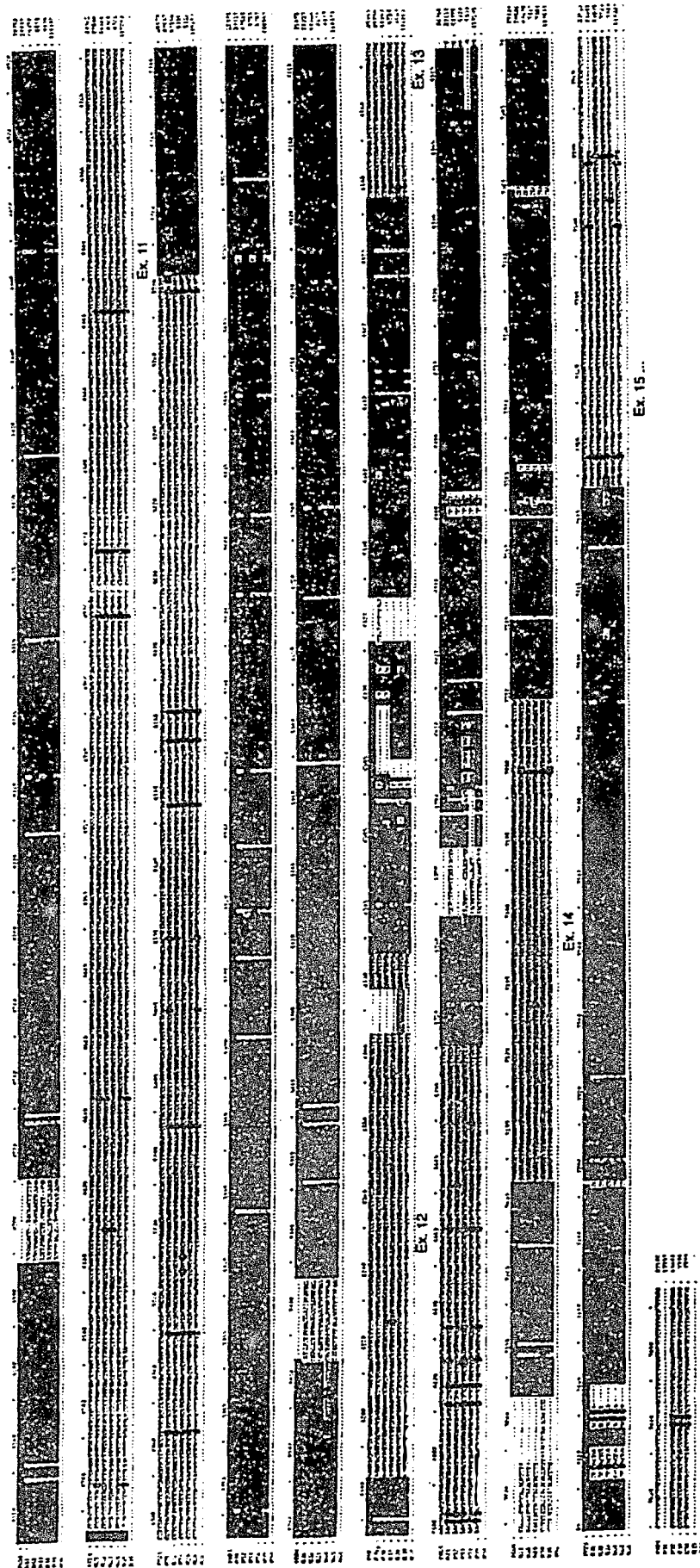


FIG. 1—Continued



TABLE 2

Degrees of Homology between PKD1 and HG in the Duplicated Region of PKD1\*

Ident	PKD1P1	PKD1P2	PKD1P3	PKD1P4	PKD1P5
PKD1	95.4%	96.7%	96.4%	96.5%	95.5%
PKD1P1	—	97.4%	97.6%	96.5%	99.9%
PKD1P2	97.4%	—	98.2%	98.0%	97.5%
PKD1P3	97.6%	98.2%	—	97.6%	97.6%
PKD1P4	96.5%	98.0%	97.6%	—	96.5%
PKD1P5	99.9%	97.5%	97.6%	96.5%	—

\* Genomic position 19866–27595.

HG1 and HG5 possess the highest degree of identity (99.9%), followed by PKD1P2 and PKD1P3 (98.2%), whereas PKD1P1 and PKD1P4 (or PKD1P5 and PKD1P4) share only 96.5% identical sequence over the examined region. On a first glance one might remain with the impression that PKD1P1 and PKD1P5 are the same gene; however, when looking closer into sequence details (Fig. 1) it is visible that these are two different genes. PKD1P1 for example should have a different splicing at the exon 11 to exon 12 junction, and PKD1P5 has other than the PKD1P1-specific deletions/insertions in intron 12, by which PKD1P5 looks more like the other HG. Overall for the region examined, PKD1 exons 2 through 14, the splicing pattern seems to be conserved for the homologous genes, with small exceptions: exon 8 for PKD1P3 and PKD1P4, the aforementioned exon 11 in PKD1P1, and exon 15 in PKD1P3. PKD1P1 and PKD1P5 share another peculiarity, 91 bp from the beginning of exon 15, there is a 4.3-kb deletion first described by Loftus *et al.* (1999), which contains most of PKD1 exon 15 and exon 16. PKD1P2, PKD1P3, and PKD1P4 do not have this feature, the PKD1 homologous sequence in PKD1P6 extends until the last 55 bp of exon 7. Most of the sequence variations are common for the HG; there are, however, many specific differences in the coding and noncoding sequences. Interestingly, exons 2, 3, and 4 and the adjacent intronic regions contain a very few variations from the PKD1 sequence. This is true for exon 9, for example, whereas exon 8 is a hot spot for

\*PKD1: AGGTTGTACAAC \*PKD1P1,3,4,5: AGGTATAGAAC \*PKD1P2:AGGTTTAGAAC

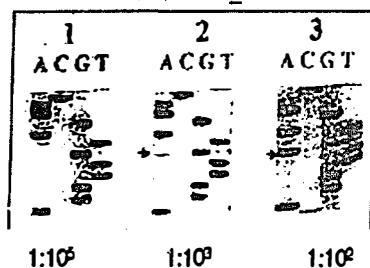


FIG. 2. Coamplification of the homologous genes, by lower dilutions of the first-round PCR. Intron 8 of PKD1 is specifically amplified in the 1:10<sup>5</sup> panel. The observed differences with HG are in position 22770 of the PKD1 genomic sequence. Note that only PKD1P2 has an "A" at this position (sequence is of the opposite strand), which disappears in the 10<sup>-3</sup> dilution, whereas the "T" (A in the opposite strand) present in four of the HG at matching location is still visible.

1 2 3 4 M

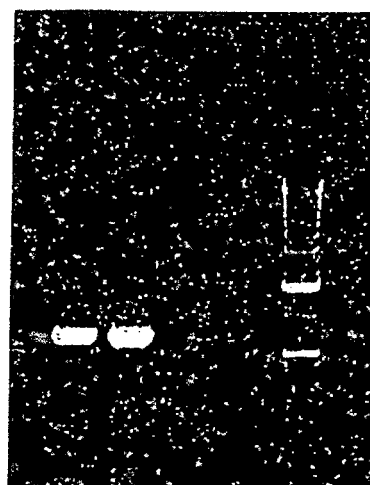


FIG. 3. RT-PCR on total RNA and translationally active mRNA (polysomal fraction) from the T98G glioblastoma cell line. Lanes 1 and 2, total RNA; lanes 3 and 4, mRNA. Lanes 1 and 3, PKD1-specific amplification; lanes 2 and 4, HG-specific amplification. M, 1 kb ladder.

such differences, six on a homologous gene on average, in a segment of 116 bp. Almost all HG, with the exception of PKD1P2, have a stop codon in exon 8 (CAG<sub>PKD1</sub> → TAG<sub>HG</sub>) with C → T at position 22690 of the PKD1 genomic sequence.

In the 7.7-kb sequenced region, two mammalian wide interspersed repeats were identified in all homologous genes and in PKD1 in matching locations (PKD1 positions 22068–22176 and 25717–25816) and PKD1P2 and PKD1 contain an additional simple repeat of a slightly different length in a matching position (PKD1 position 26133–26172). The level of repeated sequence for all genes but PKD1 and PKD1P2 is 2.7% over the analyzed segment; for PKD1P2 it is 3.7% and PKD1 has 3.2% due to a shorter simple repeat.

We used some of the identified clustered basepair differences with HG (Fig. 1) to design PKD1-specific primers for amplification of the genomic region containing exons 6 to 14 in two overlapping PCR formats and for differential amplification of exon 12 from PKD1 or HG from cDNA (see Materials and Methods). First- and second-round amplicons from human genomic DNA and T98G cDNA were also sequenced and compared with the PKD1 and HG sequences for the amplified segments. We showed that an input of a 10<sup>-6</sup> minimal dilution of the first-round PCR product into the second (nested) PCR yields a PKD1-specific product, whereas a dilution of 10<sup>-3</sup> or lower gives rise to a coamplification of the homologous genes (Fig. 2).

#### Analysis of HG Transcripts

To analyze HG transcripts further, we used RT-PCR on total RNA and translationally active mRNA from the cell line T98G with primers specific for PKD1 or for homologous genes (see Materials and Methods). We were able to amplify exon 12 of PKD1 and HG from

total RNA, as is shown on Fig. 3, lanes 1 and 2, and of PKD1 only from translationally active mRNA (lane 3). The expected PCR product of 739 bp is detected in the first three lanes, indicating that there is no genomic DNA contamination in the cDNA preparations that would result in coamplification of the abridging introns and hence in an additional band of larger size (1.8 kb). The EtBr staining of the HG amplification product is at least twice as intensive as the staining of the PKD1 amplification product by equal aliquots of the PCR reactions loaded on the gel. This could be indicative of at least two homologous genes, provided their transcription rates do not differ significantly. PKD1 and HG amplicons were confirmed by direct sequencing and it was demonstrated that HG amplicons contain all specific sequence changes characteristic for both PKD1P3 and PKD1P4 in the amplified region (results not shown).

In contrast, there is no amplification of HG from translationally active mRNA (see Materials and Methods), whereas PKD1 is amplified (Fig. 3, lanes 3 and 4). In lane 4 (HG amplification) the primers not utilized in the course of the PCR can be seen as a "primer cloud," a diffuse band of a very low size, and there is no 739-bp specific product. This failure to amplify HG from translationally active mRNA shows that HG transcripts are not bound on polysomes but rather segregated by proteins comprising the translation machinery. In addition, translation of the HG sequences we obtained in the three open reading frames (ORF) shows the presence of multiple early stop codons; in the reading frame matching the PKD1 ORF they occur already in exon 8.

## DISCUSSION

Since the discovery of genes homologous to PKD1 in 1994, many ADPKD researchers have raised the question about the possible number of HG and their eventual functional importance. The possibility that HG might code for polycystin 1-related proteins has been a major concern in localization studies on cells and tissues for the past years. Since the availability of sequence information on at least two of the homologues (Loftus *et al.*, 1999) it became evident that if translated, those genes should produce much shorter polypeptide chains of unknown functional importance compared to intact polycystin 1. The structural features of the two new HG we identified in the current study conform to this rule, indicating that premature stop codons can occur already in exon 8.

Our work confirms transcription of the homologous genes shown by others (The European Polycystic Kidney Disease Consortium, 1994). However, the new finding that HG are not translated in the model cell line T98G, in which polycystins are well expressed, guides us to the conclusion that this is the case in other cells and tissues as well, since translation of functional mRNA is a universal biological process. The polysomal fraction that we used as a source for mRNA in our RT-PCR experiments contains membrane-bound poly-

somes, some 80S ribosomes and 60S subunits, as well as mitochondrial ribosomes. The mRNA species obtained from this fraction are only those which are bound on polysomes, i.e., this is the translationally active mRNA. Homologues to PKD1 seem to produce nontranslatable mRNA species that are segregated by polysome complexes. The finding that certain mRNAs are not used as protein templates is not unusual. For example, genes in the Prader-Willi Angelman locus, PAR-SN, PAR-5, PAR-7, IPW, PAR-1, and PAR-4, produce processed but not translated mRNA species. It could as well be that the only function of those non-protein-coding transcripts is to host multiple snoRNA copies (Cavaille *et al.*, 2000). For the PKD1P locus there are no snoRNAs identified in introns. Nevertheless, translation seems not to be favored for PKD1 pseudogenes. From our analysis, almost all of them may lack the PKD1 exon 1 carrying the first ATG codon (in PKD1P1 there is a sequence gap in this region). According to the proposed scanning mechanism of translation initiation, the 40S ribosome subunit scans the transcript for an AUG codon that is in the optimal sequence context (Kozak, 1999). Downstream ATG codons in exons 6 and 9 could perhaps also make good initiation sites, since they conform to the -3 A, +4 G rule. It is, however, being demonstrated that highly structured sequences preceding a downstream AUG in the mRNA can block the advance of 40S ribosomal subunits and prevent them from reaching the start codon (Kozak, 1998). The 5' end of PKD1 and PKD1P transcripts is very rich in highly structured sequences, and it should be expected that initiation of translation cannot occur downstream of the first AUG codon in exon 1. The absence of an optimal start codon in PKD1 homologues supports the conclusion that HG are pseudogenes.

In the course of molecular evolution these sequences seem to have arisen at about the same time, judging from repeats distribution. This should be a comparatively recent event, since transcription is a very energy-demanding process, which if not bound on functional entities might not be tolerated just for pseudogene diversification. On the other hand, HG transcription could additionally regulate the level of PKD1 mRNA through recruitment and titration of transcription factors. It could be speculated that PKD1P1 and PKD1P5 are the oldest and PKD1P2 the youngest member of this pseudogene family, but additional sequencing work is necessary to support such conclusion. The exact HG number is still unknown; these genes should have arisen in a divergent fashion as paralogues to PKD1. Their repeat abundance and high GC content suggest that these pseudogenes might have emerged as a consequence of retroposition and/or nonhomologous recombination (Jurka, 1998). It is more likely that HG have resulted from recombination/duplication events, however, since they carry their introns, and the splice pattern in the PKD1 homologous regions seems to be overall conserved for these genes. Most of the retroposed sequences found in eukaryote genomes are al-

ready processed transcripts, for example the G-coupled receptor gene family and other genes resulting from retroposition of mRNA sequences (Brosius, 1999). One cannot exclude, however, even the marginal possibility that pre-mRNA sequences flanked by autonomous transposable elements could be "moved around" in the genome (Jurka, 1998). This scenario could put the accent on PKD1 and its pseudogenes as "patchworks" of minigene evolution, i.e., many small genes with defined functions assembled together to result in a new, different function. Although very attractive, such hypothesis is not confirmed by PKD1 evolutionary comparisons with the fugu fish orthologue (Sandford *et al.*, 1997). The comparative analysis has demonstrated an evolutionarily conserved sequence and an unchanged domain structure across 400 million years of vertebrate evolution. In PKD1 pseudogenes certain predicted extracellular domains of PKD1, like the leucine-rich repeat encoded by exons 2 to 4, or the exon 9 module, show extremely high degrees of conservation, in contrast to other neighboring exons. Most probably PKD1 homologues have emerged in higher primates by duplications of relevant genomic regions, since such duplicons were already identified in corresponding human chromosome 16 loci (Loftus *et al.*, 1999). Searches for PKD1-homologous sequences similar to our strategy could be applied to primate genomic libraries to find at which stage of evolution HG appeared.

The sequencing data on HG should allow for construction of better reagents for mutation analysis of PKD1 using more conventional approaches, which in turn can be applied to routine diagnosis. This information should be instructive for development of a complete PKD1 analytic system, and it can be used for a critical review of the mutation data on the duplicated part of the gene obtained by other groups.

## ACKNOWLEDGMENTS

This study was supported by grants from the Deutsche Forschungsgemeinschaft (Ho 572/18-2) and Interdisziplinäres Zentrum für Klinische Forschung, Münster (IZKF B4). The authors express their gratitude to Vladimir Kuryshov for making the sequence alignments presentable. Many thanks to Alexander Kondrashov for the interesting and fruitful discussions.

## REFERENCES

- The American PKD1 Consortium (1995). Analysis of the genomic sequence for the autosomal dominant polycystic kidney disease gene predicts the presence of a leucine-rich repeat. *Hum. Mol. Genet.* 4: 575-582.
- Bogdanova, N., Dworniczak, B., Dragova, D., Todorov, V., Dimitrakov, D., Kalinov, K., Hellmayer, J., Horst, J., and Kalaydjieva, L. (1995). Genetic heterogeneity of polycystic kidney disease in Bulgaria. *Hum. Genet.* 95: 645-650.
- Brosius, J. (1999). RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* 238: 115-134.
- Cavaille, J., Buiting, K., Kieffmann, M., Lohland, M., Brannan, C. I., Horsthemke, B., Bachelier, J. P., Brosius, J., and Huttenhofer, A. (2000). From the cover: Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl. Acad. Sci. USA* 97: 14311-14316.
- Dalgard, O. Z. (1957). Bilateral polycystic kidney disease of the kidneys: A follow-up of two hundred and eighty-four patients and their families. *Acta Med. Scand.* 328: 1-251.
- Daoust, M. C., Reynolds, D. M., Bichet, D. G., and Somlo, S. (1995). Evidence for a third genetic locus for autosomal dominant polycystic kidney disease. *Genomics* 25: 733-736.
- De Almeida, S., de Almeida, E., Peters, D., Pinto, J. R., Tavora, I., Lavinha, J., Breuning, M., and Prata, M. M. (1995). Autosomal dominant polycystic kidney disease: Evidence for the existence of a third locus in a Portuguese family. *Hum. Genet.* 96: 83-88.
- The European Polycystic Kidney Disease Consortium (1994). The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. *Cell* 77: 881-894.
- Hanaoka, K., Qian, F., Boletta, A., Bhunia, A., Piontek, K., Tsiokas, L., Sukhatme, V. P., Guggino, W. B., and Germino, G. G. (2000). Co-assembly of polycystin-1 and -2 produces unique cation-permeable currents. *Nature* 408: 121-128.
- Hughes, J., Ward, C. J., Peral, B., Aspinwall, R., Clark, K., San Millan, J. L., Gamble, V., and Harris, P. C. (1995). The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. *Nat. Genet.* 10: 151-160.
- The International Polycystic Kidney Disease Consortium (1995). Polycystic kidney disease: The complete structure of the PKD1 gene and its protein. *Cell* 81: 289-298.
- Jurka, J. (1998). Repeats in genomic DNA: Mining and meaning. *Curr. Opin. Struct. Biol.* 8: 333-337.
- Koptides, M., Hadjimichael, C., Koupepidou, P., Pierides, A., and Constantinou Deltas, C. (1999). Germinal and somatic mutations in the PKD2 gene of renal cysts in autosomal dominant polycystic kidney disease. *Hum. Mol. Genet.* 8: 509-513.
- Kozak, M. (1998). Primer-extension analysis of eukaryotic ribosome-mRNA complexes. *Nucleic Acids Res.* 26: 4853-4859.
- Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene* 234: 187-208.
- Loftus, B. J., Kim, U. J., Sneddon, V. P., Kalush, F., Brandon, R., Fuhrmann, J., Mason, T., Crosby, M. L., Barnstead, M., Cronin, L., Deslattes Mays, A., Cao, Y., Xu, R. X., Kang, H. L., Mitchell, S., Eichler, E. E., Harris, P. C., Venter, J. C., and Adams, M. D. (1999). Genome duplications and other features in 12 Mb of DNA sequence from human chromosome 16p and 16q. *Genomics* 60: 295-308.
- Mechler, B. M. (1987). Isolation of messenger RNA from membrane-bound polysomes. *Methods Enzymol.* 152: 241-248.
- Mochizuki, T., Wu, G., Hayashi, T., Xenophontos, S. L., Veldhuisen, B., Saria, J. J., Reynolds, D. M., Cai, Y., Gabow, P. A., Pierides, A., Kimberling, W. J., Breuning, M. H., Deltas, C. C., Peters, D. J., and Somlo, S. (1996). PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. *Science* 272: 1339-1342.
- Olsson, P. G., Lohning, C., Horsley, S., Kearney, L., Harris, P. C., and Frischauf, A. (1996). The mouse homologue of the polycystic kidney disease gene (Pkd1) is a single-copy gene. *Genomics* 34: 233-235.
- Pennekamp, P., Bogdanova, N., Wilda, M., Markoff, A., Hameister, H., Horst, J., and Dworniczak, B. (1998). Characterization of the murine polycystic kidney disease (Pkd2) gene. *Mamm. Genome* 9: 749-752.
- Peters, D. J. M., and Sandkuijl, L. A. (1992). Genetic heterogeneity of polycystic kidney disease in Europe. In "Contributions to Nephrology 97: Polycystic Kidney Disease" (M. H. Breuning, M. Devoto, and G. Romeo, Eds.), pp. 128-139, Karger, Basel.

- Pluthero, F. G. (1993). Rapid purification of high-activity Taq DNA polymerase. *Nucleic Acids Res.* 21: 4850-4851.
- Qian, F., Watnick, T. J., Onuchic, L. F., and Germino, G. G. (1996). The molecular basis of focal cyst formation in human autosomal dominant polycystic kidney disease type I. *Cell* 87: 979-987.
- Qian, F., Germino, F. J., Cai, Y., Zhang, X., Somlo, S., and Germino, G. G. (1997). PKD1 interacts with PKD2 through a probable coiled-coil domain. *Nat. Genet.* 16: 179-183.
- Sandford, R., Sgotto, B., Aparicio, S., Brenner, S., Vaudin, M., Wilson, R. K., Chisoe, S., Pepin K., Bateman, A., Chothia, C., Hughes, J., and Harris, P. (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.
- Tsiokas, L., Kim, E., Arnould, T., Sukhatme, V. P., and Walz, (1997). Homo- and heterodimeric interactions between the products of PKD1 and PKD2. *Proc. Natl. Acad. Sci. USA* 94: 6965-6970.
- Watnick, T. J., Torres, V. E., Gandolph, M. A., Qian, F., Onuchic, L. F., Klinger, K. W., Landes, G., and Germino, G. G. (1997). Somatic mutation in individual liver cysts supports a two model of cystogenesis in autosomal dominant polycystic kidney disease. *Mol. Cell* 2: 247-251.

### **Appendix B3**

McCluskey *et al.* (2002). Mutation detection in the duplicated region of the polycystic kidney disease 1 (PKD1) gene in PKD1-linked Australian families. *Hum Mutat.* 19(3): 240-250.

RESEARCH ARTICLE

# Mutation Detection in the Duplicated Region of the Polycystic Kidney Disease 1 (*PKD1*) Gene in PKD1-Linked Australian Families

Marie McCluskey,<sup>1</sup> Tina Schiavello,<sup>1</sup> Michael Hunter,<sup>1</sup> Janina Hantke,<sup>1,4</sup> Dora Angelicheva,<sup>1</sup> Nadja Bogdanova,<sup>2</sup> Arseni Markoff,<sup>2</sup> Mark Thomas,<sup>3</sup> Bernd Dworniczak,<sup>2</sup> Juergen Horst,<sup>2</sup> and Luba Kalaydjieva<sup>1,4\*</sup>

<sup>1</sup>Centre for Human Genetics, Edith Cowan University, Joondalup, Australia

<sup>2</sup>Institut für Humangenetik, Westfälische Wilhelms-Universität, Münster, Germany

<sup>3</sup>Department of Nephrology, Royal Perth Hospital, Perth, Australia

<sup>4</sup>WAIMR, West Australian Institute of Medical Research, Perth, Australia

Communicated by Peter Pearson

Screening for disease-causing mutations in the duplicated region of the *PKD1* gene was performed in 17 unrelated Australian individuals with PKD1-linked autosomal dominant polycystic kidney disease. Exons 2–21 and 23–34 were assayed using PKD1-specific PCR amplification and direct sequencing. We have identified 12 novel probably pathogenic DNA variants, including five truncating mutations (Q563X, c.5105delAT, c.5159delG, S2269X, c.9847delC), two in-frame deletions (c.7472del3, c.9292del39), and two splice-site mutations (IVS14+1G>C, IVS16+1G>T). Three of the mutations (G381C, Y2185D, G2785D) were predicted to lead to the replacement of conserved amino acid residues, with ensuing changes in protein conformation. Defects in the duplicated region of PKD1 thus account for 63% of our patients. Together with the previously detected mutations (Q4041X, R4227P) in the 3' region of the gene, the study has achieved an overall mutation detection rate of 74%. In addition, we have detected 31 variants (nine novel and 22 previously published) that did not segregate with the disease and were considered to be neutral polymorphisms. Three of the nine novel polymorphisms were missense mutations with a predicted effect on protein conformation, emphasizing the problems of interpretation in PKD1 mutation screening. Hum Mutat 19:240–250, 2002. © 2002 Wiley-Liss, Inc.

**KEY WORDS:** PKD1; polycystic kidney disease; autosomal dominant; ADPKD; polycystin-1; genotype-phenotype; homologous genes; pseudogenes; SNP

**DATABASES:**

PKD1 – OMIM:601313, 173900 (PKD); GDB:120293; GenBank:L39891 (PKD1); AC002039, AF320593, AF320594, AC010488, AC040158 (PKD1-homologous genes)

## INTRODUCTION

Autosomal dominant polycystic kidney disease (ADPKD; MIM# 173900) is a common monogenic disorder with an incidence of approximately one in 1,000 [Daugaard, 1957]. The disease manifests with the development of bilateral kidney cysts followed by renal failure in most affected individuals. Eight to 10% of all patients requiring hemodialysis or renal transplantation have ADPKD [Gabow, 1993]. Additional symptoms, such as hypertension, urinary tract infections, hematuria, proteinuria, extra-renal cysts,

cerebral aneurysms, herniae, cardiac valve defects, renal calculi, and gastrointestinal diverticula, vary both between and within families [Roscoe et al., 1993; Schievink and Torres, 1997]. The rate of progression toward kidney failure is also variable: end-stage renal failure

Received 1 June 2001; accepted revised manuscript 19 November 2001.

\*Correspondence to: Luba Kalaydjieva, Centre for Human Genetics, Edith Cowan University, Joondalup, WA 6027, Australia. E-mail: l.kalaydjieva@ecu.edu.au

(ESRF) occurs most commonly in the fifth decade [Milutinovic et al., 1984]. However, it can present as early as in utero, while some individuals may never progress to renal failure [Choukroun et al., 1995; Parfrey et al., 1990; Perrichot et al., 2001]. The wide range of clinical manifestations and severity is probably a compound product of heterogeneous germline mutations, and of the random nature of the "second hit," namely somatic mutations in the two characterized genes for ADPKD, PKD1 (MIM# 601313), and PKD2 (MIM# 173910) [Qian et al., 1996; Watnick et al., 1998; Koptides et al., 1999].

Up to 85% of germline mutations in Caucasian ADPKD individuals affect the *PKD1* gene [Peters and Sandkuijl, 1992], located on chromosome 16p13.3 [The European Polycystic Kidney Disease Consortium, 1994]. The *PKD1* gene consists of 46 exons spanning ~52 kb of genomic DNA [The International Polycystic Kidney Disease Consortium, 1995; The American PKD1 Consortium, 1995; Hughes et al., 1995]. A large part of the gene is duplicated in at least six homologous genes (HGs), which show > 95% homology with the *PKD1* gene [Hughes et al., 1995; Bogdanova et al., 2001]. The homologous genes cause problems of co-amplification, thus creating major obstacles to the mutation analysis of patients and families with the most common form of ADPKD.

Here we present the results of a screening for mutations in the duplicated region of *PKD1* in 17 Western Australian ADPKD families. Using *PKD1*-specific oligonucleotide primers and direct automated DNA sequencing, we have identified 12 putative pathogenic mutations and achieved an overall *PKD1* mutation detection rate of 74%.

## SUBJECTS AND METHODS

### Subjects

Nineteen ADPKD families, 18 of British and one of Spanish descent, were recruited through index patients at the Department of Nephrology, Royal Perth Hospital, Western Australia, as part of a large collaborative study of ADPKD. The diagnosis was based on standard ultrasound criteria [Bear et al., 1992]. Preliminary analysis of these families had shown linkage to the *PKD1* gene. Two families had mutations in the unique part of the *PKD1* gene [Bogdanova et al., 2000]. The screening for mutations in the duplicated

region of *PKD1* included the remaining 17 affected individuals.

### DNA Isolation and PCR Amplification

DNA was extracted from whole blood using standard procedures [Miller et al., 1988]. Primers for long-range amplification were chosen to be specific to *PKD1* when compared with its homologous genes. All long-range polymerase chain reaction (XL-PCR) primers are specific to *PKD1* when compared to HG1 and HG2 [Loftus et al., 1999]. The *PKD1* sequence between exons 2 to 14 was compared also to HG3, HG4, and HG5 [Bogdanova et al., 2001]. All primers in this region are specific to *PKD1* when compared to HG3 and HG5. Our XL-PCR products are not specific to *PKD1* when compared with HG4 between exons 6 and 11, due to the identity between the two sequences. However, the nested primers used for amplification of exons 6, 9, 10, and 11 (fragment A) are specific to *PKD1* when compared to HG4.

XL-PCR was performed on 400 ng of genomic DNA in a final volume of 100  $\mu$ l using 4 units of Applied Biosystems XL-rTth Taq polymerase with 200  $\mu$ M of each dNTP, 120 ng each of forward and reverse primer, 1.5 mM MgOAc, and 1  $\times$  buffer (Applied Biosystems, Foster City, CA). The optimized conditions for XL-PCR are listed in Table 1.

XL-PCR products (5  $\mu$ l aliquots) were electrophoresed in 1% agarose gels and visualized with ethidium bromide staining in order to confirm an amplification product of the expected length and to rule out large deletions or insertions.

Nested PCR reactions utilized 1  $\mu$ l of XL-PCR template diluted by a factor of at least  $1 \times 10^{-4}$  to prevent co-amplification of the *PKD1* homologous genes [Watnick et al., 1997; Bogdanova et al., 2001]. Nested PCR was performed in a final volume of 25  $\mu$ l with 0.2 units of QIAGEN<sup>®</sup> Taq DNA polymerase with 200  $\mu$ M of each dNTP, 60 ng each of forward and reverse primer, 20% Q-solution (QIAGEN, Hilden, Germany), 1.5 mM MgCl<sub>2</sub>, and 1  $\times$  buffer (QIAGEN). The optimized conditions and specific oligonucleotide primers for nested PCR amplification of individual fragments are listed in Table 2.

### DNA Sequencing

All samples were subjected to cycle sequencing in both directions with the BigDye<sup>™</sup> Termini-

TABLE 1. Primers for XL-Template Amplification

Exons	Primer sequence	Cycling conditions	Genomic position	Product size (bp)
Ex 2-5	IVS1F 5' ccagcgtctcatctgtctgg 3' <sup>a</sup>	68°C (X 40)	17049-69	4,542
XL-1	IVS5R 5' ggtgggaacgagggtgtcac 3'	2-step (0.5/5')	21579-90	
Ex 6-11	2927 5' gccgctacagcatcgtggccc 3'	68°C (X 40)	21447-470	4,198
XL-2	11R 5' cagtggaggaggacacactact 3'	2-step (0.5/4')	25622-644	
Ex 12-14	12F 5' cagtgtagggtgcctccctactg 3'	68°C (X 40)	25621-644	1,810
XL-3	3130 5' ccagcacggtcaggagggtact 3'	2-step (0.5/2')	27410-430	
Ex 15	IVS13F 5' ctgtcccggtctcactcctgc 3'	68°C (X 40)	26769-89	3,861
XL-4	Ex15XLR 5' caacgtgggctccaagtagtg 3'	2-step (0.5/4')	30607-29	
Ex 15-21	Ex15XLF 5' gcaactacttgaggcccaag 3'	68°C (X 40)	30606-26	3,375
XL-5	IVS21R 5' gcagggtgagcagggtggggccac 3'	2-step (0.5/4')	33957-80	
Ex 23-24	Ex23F 5' ctgcactgacccacgcattg 3' <sup>b</sup>	68°C (X 40)	37674-694	6,749
XL-6	Ex34R 5' atgtgggtgtcttggtagg 3'	2-step (0.5/6')	44403-422	

<sup>a</sup>IVS1F used by Thomas et al. [1999] designated 17EXT.

<sup>b</sup>Ex23F used by Watnick et al. [1997] designated TWF1.

nator Cycle Sequencing Ready Reaction Kit (Applied Biosystems), using the nested PCR primers listed in Table 2. The products were electrophoresed on an automated 377 DNA Analyzer (Applied Biosystems).

Deletions and insertions were characterized by performing a fresh PCR reaction, cloning the products using the TOPO™ TA Cloning® Kits with the pCR2.1-TOPO™ vector (Invitrogen, San Diego, CA), and subsequent sequencing.

#### Data Analysis

Sequences were analyzed using Sequence Navigator 1.0.1 (Applied Biosystems) and Sequencher 3.0 (Gene Codes Corporation, Ann Arbor, MI). Missense mutations were analyzed further for the conservation of the amino acid residue in the mouse and pufferfish protein sequences, using Pfam ([www.sanger.ac.uk/Software/Pfam/](http://www.sanger.ac.uk/Software/Pfam/)). PHDsec [Rost and Sander, 1994] was used to predict the effect of the mutation on secondary protein structure.

Confirmation of putative pathogenic sequence variants was achieved through repeated analysis of the index patient and a segregation study including affected and unaffected family members. Segregation analysis was not possible in one out of 12 families, where linkage analysis had been performed on DNA extracted from blood spots collected on Guthrie cards.

Mutations are named following the recommendations of Cotton and Scriver [1998], whereas nucleotide positions are numbered in the traditional way, where the first coding nucleotide of the cDNA is at position 1212 for genomic sequence Accession # L33243 and position 3648 for Accession # L39891.

## RESULTS

Using the techniques outlined above, we have identified a total of 43 DNA sequence variations in the duplicated region of the PKD1 gene in the 17 unrelated affected individuals. The detected variants include 12 probable pathogenic mutations, all of which are reported for the first time (Table 3 and Fig. 1), and 31 neutral polymorphisms, including nine novel and 22 previously reported (Table 4).

#### Probable Disease-Causing Mutations

**Protein truncating mutations.** Two nonsense and three frameshift mutations were detected in the study.

**Q563X.** The C>T transition at c.1898 (g.22716) in exon 8 is predicted to truncate polycystin-1 by 3,739 amino acids, corresponding to the loss of most of the protein.

**S2269X.** S2269X results from a C>G transversion at c.7017 (g.30915) in exon 15. It is predicted to truncate polycystin-1 by 2,033 amino acids, spanning the REJ domain and all of the transmembrane and cytoplasmic regions of the protein.

**c.5105-5106delAT.** A 2 bp deletion, c.5105-5106delAT (g.29004-29005delAT), was found in exon 15 of the PKD1 gene. The mutation causes a frameshift at amino acid position 1632 and a premature termination signal at codon 1656, corresponding to the loss of most of the PKD repeats, the REJ domain, the transmembrane domains, and the entire cytoplasmic region of the protein.

**c.5159delG.** A single bp deletion, c.5159delG (g.29058delG), was detected in exon 15 of the gene, leading to a shift in the reading frame at



TABLE 2. Primers and Conditions for Nested PCR Amplification of PKD1

Exons	Primer sequence	Cycling conditions	Genomic position	Product size (bp)
Ex 2	Ex2F 5' ggaaggccctdggcctcta 3'	64°C (X 35)	19845-64	230
XL-1	Ex2R 5' ggggaatcagcaagctgat 3'	(0.5/0.5/1')	20064-74	
Ex 3	Ex3F 5' ctacgtgtggggattcca 3'	58°C (X 35)	20045-64	186
XL-1	Ex3R 5' ggcagaaggatattggggg 3'	(0.5/1/1')	20211-30	
Ex4	Ex4F 5' ggcctggcatagaccctccc 3'	64°C (X 35)	20378-97	273
XL-1	Ex4R 5' cctggctgggaaggacagag 3'	(0.5/0.5/1')	20630-50	
Ex 5(a)	Ex5aF 5' gggagccaggaggagca 3'	58°C (X 35)	20770-87	329
XL-1	Ex5aR 5' ggggaagcgtgctggagg 3'	(0.5/1/1')	21081-98	
Ex 5(b)	Ex5bF 5' cctccagcagctctccc 3'	64°C (X 35)	21081-98	272
XL-1	Ex5bR 5' ccgctccacctgcagct 3'	(0.5/0.5/1')	21335-52	
Ex 5(c)	Ex5cF 5' ctacgctgctggggacaga 3'	64°C (X 35)	21315-35	275
XL-1	IVS5 R 5' ggtgggaacgagggtgtcaac 3'	(0.5/0.5/1')	21579-89	
Ex 6	4349 5' gtgagtgtctgctgccca 3'	58°C (X 30)	21491-508	396
XL-2	4350 5' cctcctcctcctgagact 3'	(1/1/1')	21868-886	
Ex 7	4351 5' ggcctgagcctcagttt 3'	58°C (X 30)	22104-121	391
XL-2	4352 5' taaccacagccagcgtct 3'	(1/1/1')	22476-493	
Ex 8	4353 5' gctgtgtcgtcctggtgt 3'	58°C (X 30)	22575-592	306
XL-2	4354 5' ccaatctcactgggcaca 3'	(1/1/1')	22863-880	
Ex 9	4345 5' gttcggtagggggagttct 3'	60°C (X 35)	23090-108	271
XL-2	4346 5' gtgaagctcagagaggcca 3'	(0.8/0.8/0.8')	23341-360	
Ex 10	4347 5' ggtggcctgtgggcaaatca 3'	61°C (X 30)	23608-527	340
XL-2	4348 5' gccctgaggagatgcaggga 3'	(1/1/1')	23929-947	
Ex 11-11a	11AF 5' gggccacgggacatgacagt 3'	66°C (X 25)	24300-320	331
XL-2	11AR 5' cgcagctccagggtgggctc 3'	(0.5/1/1')	24610-630	
Ex 11-11b	11BF 5' acggaaacagctcacctgct 3'	66°C (X 25)	24581-600	320
XL-2	11BR 5' aaggtggccaccaggcgagg 3'	(0.5/1/1')	24880-900	
Ex 11-11c	11CF 5' gcaagtgtcagcgcctgctt 3'	66°C (X 25)	24849-869	321
XL-2	11CR 5' agcaccctgtcgcaggcac 3'	(0.5/1/1')	25150-170	
Ex 12	4355 5' ccaggaggcagacagccta 3'	61°C (X 30)	25923-940	352
XL-3	5135 5' cctggccctgattggcgtc 3'	(0.8/0.8/0.8')	26256-74	
Ex 13	5228 5' tcaactgccacctggctcac 3'	68°C (X 35)	26279-299	320
XL-3	5136 5' ccgaggctcagaaagca 3'	2-step (0.5/2')	26581-598	
Ex 14	5002 5' ctgtcccggttcactactgc 3'	68°C (X 35)	26769-789	275
XL-3	4140 5' gagggtgtgtggggaggaagg 3'	2-step (0.5/2')	27023-043	
Ex 15	Amplified as 16 fragments using primers and conditions described previously by Wainick et al. [1999]			
XL-4				
Ex 15-15(a)	Ex15LF 5' gcaactacttggaggccacg 3'	66°C (X 35)	30606-26	287
XL-5	Ex15aR 5' ctcaatgaggcaccaggcg 3'	(0.5/0.5/1')	30872-92	
Ex 15-15(b)	Ex15bF 5' catccaggccatgtgacggg 3'	64°C (X 35)	30841-61	266
XL-5	Ex15bR 5' cctgtggcaggctgggtgt 3'	(0.5/0.5/1')	31086-106	
Ex 16	Ex16F 5' ctggatggggctctcagg 3'	56°C (X 30)	31197-214	289
XL-5	Ex16R 5' cctgtggcaggctgggtgt 3'	(0.5/0.5/1')	31468-85	
Ex 17	Ex17F 5' gggctccctcagctctccag 3'	66°C (X 35)	32262-81	247
XL-5	Ex17R 5' ccaatccacagcccgccaca 3'	(0.5/0.5/1')	32489-508	
Ex 18	Ex18F 5' gctccctcaccacccctct 3'	66°C (X 35)	32571-90	344
XL-5	Ex18R 5' gatcccgctgcctcccca 3'	(0.5/0.5/1')	32895-914	
Ex 19	Ex19F 5' gtgatccgtggggacgctc 3'	66°C (X 35)	32946-65	287
XL-5	Ex19R 5' gtgagcagggtggcagctcg 3'	(0.5/0.5/1')	33213-32	
Ex 20	Ex20F 5' ctgctaccacccctctg 3'	64°C (X 35)	33225-43	239
XL-5	Ex20R 5' ggtcccaagcacgagtgca 3'	(0.5/0.5/1')	33445-63	
Ex 21	Ex21F 5' cgtctgtgacagctgtgtgcc 3'	66°C (X 30)	33763-86	218
XL-5	IVS21R 5' gcagggtgagcaggtggggccatc 3'	0.5/0.5/1')	33957-80	
Ex 23-23a	23aF 5' ctgcactgacacacagctgt 3'	58°C (X 30)	37674-694	353
XL-6	23aR 5' caccagaagatgggctgcac 3'	(0.5/0.5/1')	38006-026	
Ex 23-23b	23bF 5' ggggcccctggcacaactcag 3'	68°C (X 35)	37979-998	377
XL-6	23bR 5' cgtgtgcccacccgctgca 3'	2-step (0.5/2')	38336-355	
Ex 24	24F 5' tgtgacctgcgctctg 3'	68°C (X 35)	38600-617	204
XL-6	24R 5' ccaggctggcccgagag 3'	2-step (0.5/2')	38786-803	
Ex 25	25F 5' ctgggtcactgctccgctacc 3'	68°C (X 35)	38927-946	329
XL-6	25R 5' cctcgactcagagggctc 3'	2-step (0.5/2')	39236-255	
Ex 26	26F 5' cggctctactcagagagc 3'	68°C (X 35)	39272-291	315
XL-6	26R 5' cagcacagccagtgcagca 3'	2-step (0.5/2')	39567-586	

(Continued)

TABLE 2. Continued.

Exons	Primer sequence	Cycling conditions	Genomic position	Product size (bp)
Ex 27	27F 5' cctccacccctccctctg 3'	68°C (X 35)	41000-019	225
XL-6	27R 5' cagagctggcagggtccgc 3'	2-step (0.5/2')	41205-224	
Ex 28	28F 5' cagagctgacctccctctg 3'	58°C (X 30)	41254-273	204
XL-6	28R 5' cggagtgaggacctggacg 3'	(0.5/0.5/1')	41438-457	
Ex 29	29F 5' cgtggcctctgcagtcggg 3'	68°C (X 35)	41486-505	285
XL-6	29R 5' ggcagggaaggaggtgcccgc 3'	2-step (0.5/2')	41751-770	
Ex 30	30F 5' caccctgtgtggcctctctc 3'	68°C (X 35)	41794-813	199
XL-6	30R 5' ggctccattcccagttactcc 3'	2-step (0.5/2')	41973-992	
Ex 31	31F 5' gctgacctgacctctgc 3'	68°C (X 35)	43585-603	207
XL-6	31R 5' agtccaagctgcgcaagg 3'	2-step (0.5/2')	43773-791	
Ex 32	32F 5' ttggcgagcttgagctc 3'	58°C (X 30)	43775-792	123
XL-6	32R 5' cagggtctgaggtttctc 3'	(0.5/0.5/1')	43879-896	
Ex 33	33F 5' taccctgtgacctccgc 3'	68°C (X 35)	44064-081	250
XL-6	33R 5' ggtgagctcagagcccc 3'	2-step (0.5/2')	44296-313	

codon 1650 and a termination signal at position 1721. The mutation is predicted to truncate polycystin-1 by 2,581 amino acids, including the domains described in the above paragraph.

**c.9847delC.** A single bp deletion in exon 28, c.9847delC (g.41352delC), causes a frameshift at position 3213 and generates a premature stop codon at residue 3315. Polycystin-1 is truncated by 987 amino acids, corresponding to most of its cytoplasmic region.

**Splicing mutations.** We have identified two nucleotide substitutions affecting donor splice sites and therefore predicted to result in aberrant RNA splicing.

**IVS14+1G>C.** A G>C transversion at g.26942, 1 bp downstream of exon 14, abolishes the consensus donor splice signal in IVS 14.

**IVS16+1G>T.** A G>T transversion at g.31394G>T, 1bp downstream of exon 16, affects the consensus sequence of the donor splice site in IVS 16.

**In-frame deletions and missense mutations.** These mutations were classified as probably pathogenic on the basis of family segregation analysis, the evolutionary conservation of the amino acid residues involved and/or the predicted changes in the secondary structure of the protein.

**c.7472-7474del3.** A three base pair deletion in exon 18, c.7472-7474delACC (g.32650-32652del), results in the deletion of threonine at codon 2421 in the REJ domain of polycystin-1. This is the second threonine in a group of three that are conserved in the mouse and *Fugu* proteins.

TABLE 3. Summary of Novel, Probable Disease-Causing Mutations in the 5' Duplicated Region of PKD1

Location	Sequence variation	Amino acid change (and protein region or domain)	Predicted effect on protein	Affected family	Family segregation
Ex 8	c.1898C>T g.22716C>T	Q563X (Extracellular)	Truncation (nonsense)	A4	Yes
Ex 15	c.7017C>G g.30915C>G	S2269X (REJ domain)	Truncation (nonsense)	A15	Yes
Ex 15	c.5105-5106del (AT) g.29004delAT	I1632fsX1656 (PKD domain)	Truncation (frame-shift deletion)	A52	Not possible
Ex 15	c.5159delG g.29058delG	V1650fsX1721 (PKD domain)	Truncation (frame-shift deletion)	A36	Yes
Ex 28	c.9847delC g.41352delC	F321fsX3315 (PLAT domain)	Truncation (frame-shift deletion)	A10	Yes
IVS 14	IVS14+1G>C g.26942G>C	Intronic	Aberrant splicing	A22	Yes
IVS 16	IVS16+1G>T g.31394G>T	Intronic	Aberrant splicing	A5	Yes
Ex 5	c.1352G>T g.21430G>T	G381C (Extracellular)	Missense (loop and helix shortened)	A9	Yes
Ex 15	c.6764T>G g.30662T>G	Y2185D (PKD domain)	Missense (loop lost)	A61	Yes
Ex 23	c.8565G>A g.37893G>A	G2785D (Extracellular)	Missense (loop slightly shortened)	A6	Yes
Ex 18	c.7472-74del (ACC) g.32650-52del	2421delT (REJ domain)	In-frame deletion (helix replaced by extended sheet)	A46	Yes
Ex 25	c.9292-9331del g.39090del39	3027del13 aa (Extracellular)	In-frame deletion (loop lost)	A42	Yes

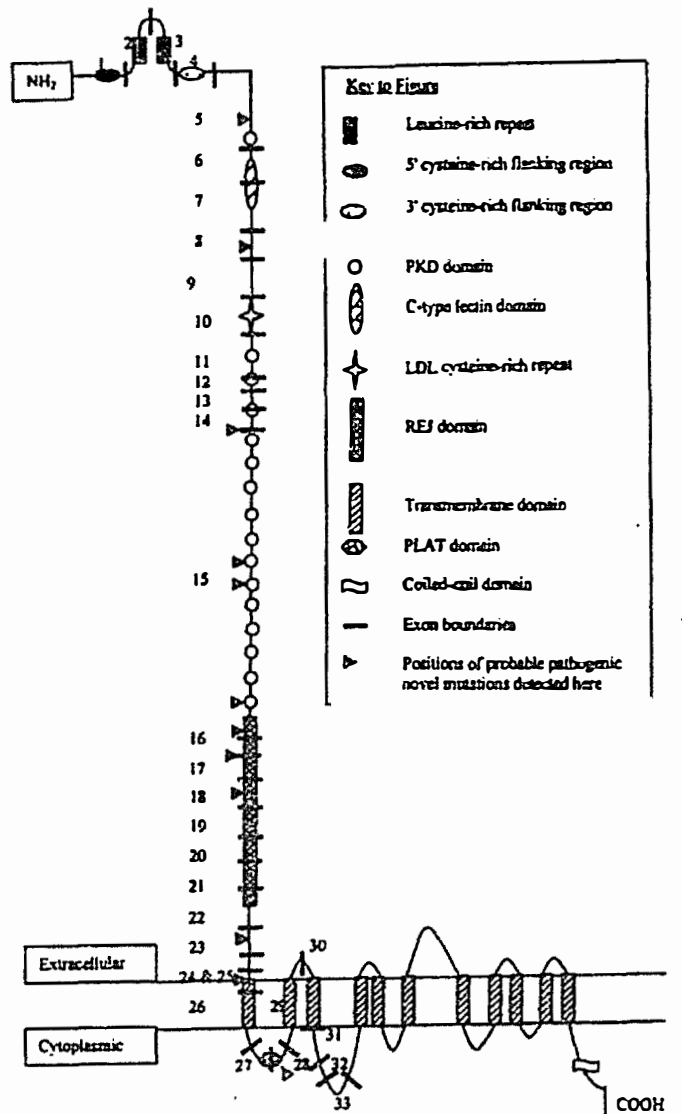


FIGURE 1. Schematic representation of polycystin-1 [adapted from Phakdeekittcharoen et al., 2000]. The locations of mutations detected in this study are marked with an arrow.

**c.9292-9331del.** This 39 base pair deletion in exon 25 is expected to result in the deletion of 13 highly conserved amino acid residues (positions 3027 to 3039) preceding the first transmembrane domain of polycystin-1. An example of the predicted topology change for this deletion is shown in Figure 2.

**G381C.** A G>T transversion at c.1352 (g.21430) in exon 5 leads to the replacement of a conserved glycine by a cysteine residue at position 381, between the first PKD domain and the c-type lectin domain.

**Y2185D.** A T>G transversion at c.6764 (g.30662) in exon 15 results in the substitution of histidine for tryptophan at the conserved position 2185 in the PKD repeat region of polycystin-1. An example of the evolutionary conservation and topology prediction for this missense change is shown in Figure 2.

**G2785D.** A G>A transition at c.8565 (g.37893) in exon 23 is predicted to result in a histidine for glycine substitution at conserved position 2785, immediately downstream of the REJ domain.

TABLE 4. Polymorphisms in the 5' Duplicated Region of PKD1

Location	Number of patients with change	Sequence variation	Amino acid position	Type of change	Previous reference
Exon 2	1	c.487G>A g.19963G>A	A92A	Silent exonc	Rossetti et al. [2001]
Exon 5	1	c.1234C>T g.21312C>T	A341A	Silent exonc	Thomas et al. [1999]
Exon 5	6 (A15.1)*	c.1330T>C g.21408T>C	L373L	Silent exonc	Rossetti et al. [2001]
IVS 6	1	g.22160G>A	—	Intronic	Novel
IVS 9	3	g.23307del7 (VNTR)	—	Intronic	Novel
Exon 11	1	c.2427A g.24474G>A	Q739R	Missense	Novel
Exon 11	1	c.2905A>C g.24952A>C	A898A	Silent exonc	Thomas et al. [1999]
Exon 11	3	c.2911G>A g.24958G>A	P900P	Silent exonc	Thomas et al. [1999]
Exon 11	3	c.2941C>T g.24988C>T	D910D	Silent exonc	Thomas et al. [1999]
Exon 13	3	c.3274T>C g.26395T>C	G1021G	Silent exonc	Thomas et al. [1999]
Exon 13	1	c.3319G>T g.26440G>T	T1036T	Silent exonc	Novel
Exon 13	4	c.3322A>G g.26443A>G	L1037L	Silent exonc	Thomas et al. [1999]
Exon 14	3	c.3486T>C g.26921T>C	M1092T	Missense	Novel
Exon 15	2	c.3583C>T g.27483C>T	A1124A	Silent exonc	Thomas et al. [1999]
Exon 15	2	c.3586C>T g.27486C>T	S1125S	Silent exonc	Thomas et al. [1999]
Exon 15	2	c.4406T>C g.28306T>C	W1399R	Missense	Watnick et al. [1999]
Exon 15	5	c.4876A>C g.28775A>C	A1555A	Silent exonc	Watnick et al. [1999]
Exon 15	2	c.4885G>A g.28784G>A	T1558T	Silent exonc	Watnick et al. [1999]
Exon 15	1	c.5157C>T g.29056C>T	T1649M	Missense	Novel
Exon 15	5	c.5383C>T g.29280C>T	T1724T	Silent exonc	Watnick et al. [1999]
Exon 15	3	c.5974G>A g.29872G>A	L1921L	Silent exonc	Thomas et al. [1999]
Exon 16	1	c.7138C>T g.31255C>T	G2309G	Silent exonc	Perrichot et al. [2000b]
IVS 16	1	g.31414G>A	—	Intronic	Novel
Exon 17	4	c.7376T>C g.32427T>C	L2389L	Silent exonc	Watnick et al. [1999]
Exon 18	4	c.7652C>T g.32830C>T	L2481L	Silent exonc	Perrichot et al. [2000b]
Exon 21	1	c.8098G>A g.33824G>A	A2629A	Silent exonc	Novel
Exon 21	2	c.8124A>G g.33849A>G	H2638R	Missense	Watnick et al. [1999]
Exon 23	2	c.8504C>T g.37854C>T	R2765C	Missense	Rossetti et al. [2001]
Exon 25	5	c.9406GT>CC g.39207GT>CC	F3066L	Missense	Peral et al. [1997]
Exon 26	5	c.9541T>C g.39466T>C	P3110P	Silent exonc	Peral et al. [1997]
IVS 31	1	g.43740T>C	—	Intronic	Novel

\*A15.1 is homozygote for c.1330T>C. All other individuals are heterozygotes for polymorphisms listed.

### Probable Neutral Polymorphisms

The 31 polymorphic variants detected during our analysis of the duplicated region of PKD1 are shown in Table 4. Here we present in more detail the novel missense variants.

**Q739R.** A G to A transition (c.2427, g.24474) in exon 11 was predicted to result in an amino acid substitution at codon 739. Human sequence Accession # L33242 and mouse Pkd1 Accession # O08852 suggest a conserved glutamine at that position. However, another published human sequence, Accession #P98161, lists the same residue as arginine, suggesting a polymorphic variant. Computer prediction models show a marked difference in protein secondary structure (Fig. 2). In this study, the glutamine to arginine change was detected in a single affected subject (A42.1). The same substitution has been found subsequently in an ADPKD family from Bulgaria, where it did not segre-

gate with the disease (N. Bogdanova, unpublished results).

**M1092T.** A T to C substitution (c. 3486, g.26921) in exon 14 was found to occur in three unrelated affected subjects. Published human sequences have a methionine residue, while the mouse protein contains threonine. Again, a difference in secondary protein structure, albeit less pronounced, is predicted by computer analysis.

**T1649M.** A C to T substitution (c.5157, g.29056) in exon 15 results in the replacement of the conserved threonine at position 1649 by methionine. Computer models predict a slight change in secondary structure. This variant was detected in patient A52.1, where family segregation analysis was not possible. However, the cloning experiment performed to characterize the frameshift deletion (c.5105delAT) of the same subject demonstrated that the missense variant was carried on the other chromosome.

FIGURE 2. Evolutionary conservation and topology prediction for mutations Y2185D, and c.9292-9331del; and for polymorphism M/T 1092. Variable amino acids are shown in bold and underlined. The human and mouse (and where possible the *Fugu*) protein sequences are aligned. Predicted protein topologies for the wild-type and variant alleles are shown for each example. The abbreviations of E (extended sheet), H (helix), and L (loop) are in accordance with PHDsec [Rost and Sander, 1994].

Mutation detection in the duplicated region of the *PKD1* gene has been hindered by the high (>95%) sequence identity with the homologous genes, raising concerns about PCR primer specificity and hence the reliability of the analysis. Most studies have therefore focused on a limited area of the gene. Only a few have screened most or the entire coding region of *PKD1*

[Watnick et al., 1999; Thomas et al., 1999; Rossetti et al., 2001; Phakdeekitcharoen et al., 2001]. In this study, we have used information published previously on the sequence of HG1 and HG2 [Loftus et al., 1999], as well as our own data on additional homologous genes [Bogdanova et al., 2001], to design amplification primers that were specific to the duplicated region of PKD1. The analytical protocol in-

cluded long-range amplification, followed by nested PCR and direct sequencing.

We have tested 17 affected subjects for mutations in exons 2–21 and 23–34, which account for about 69% of the *PKD1* coding sequence. A total of 12 possible disease-causing mutations were identified, spread from exon 5 to exon 28. The results of this study and others [Krawczak and Cooper, 1997; Afzal et al., 2000; Aguiari et al., 2000; Bogdanova et al., 2000; Kim et al., 2000; Koptides et al., 2000; Perrichot et al., 2000a,b; Phakdeekitcharoen et al., 2000, 2001; Rossetti et al., 2001] suggest that mutations are dispersed across the entire *PKD1* gene, without any apparent clustering in mutation “hot spots.”

Similar to previous studies [Krawczak and Cooper, 1997; Afzal et al., 2000; Aguiari et al., 2000; Phakdeekitcharoen et al., 2000, 2001], all mutations that we have identified were private, i.e. confined to individual families.

Our analytical procedure is not problem-free: it is relatively costly and includes sequential amplification steps that can introduce sequence errors and therefore should be avoided in diagnostic protocols. However, in the case of ADPKD, the choice is limited by the characteristics of the *PKD1* gene and of the pathogenic mutations. The risk of non-specific PCR amplification, due to the homologous genes, imposes the need of consecutive amplification steps. The spread of mutations across the *PKD1* gene, their diversity (with truncating defects accounting for a little over 50% of our patients), and their private nature, all interfere with designing a hierarchical diagnostic procedure, prioritizing common defects and mutation hot spots.

The large number of private ADPKD mutations, together with the proposed random second-hit mechanism [Qian et al., 1996], make phenotype-genotype correlations difficult to assess. The analysis of such correlations will have to be limited to comparing the general effects of the nature of mutations (truncating vs. missense) and localization in specific protein domains. In our small group of patients, there was no significant difference between the age at onset of ESRF as related to the type of germline mutation (data not shown).

In view of the lack of functional assays and the prohibitively high costs of analyzing the *PKD1* gene in a sufficient number of control chromosomes, the distinction between disease-

causing missense mutations, and “neutral variants” resulting in amino acid substitutions, has to rely on family segregation studies, assessment of the evolutionary conservation of the amino acid residue involved, the nature of the substitution, and the predicted effect on protein conformation. The predictive value of segregation analysis is limited in small families, as co-inheritance can be due to chance. As for evolutionary conservation and conformational changes, all three novel missense mutations, classified as non-pathogenic in our study due to lack of cosegregation, were predicted to result in conformational changes and one of the three (T/M1649) was conserved between mouse and human, emphasizing the problems of interpretation in *PKD1* mutation analysis. Moreover, while independent segregation of missense mutations and the ADPKD phenotype indicates a lack of drastic effect on protein function, modifying effects on disease severity remain a possibility and merit further studies. In addition, the accumulation of “neutral” polymorphisms may affect protein function through numerous changes in topology even in the absence of an obvious pathogenic mutation. The existence of such “hypomorphic” alleles has been proposed by Watnick et al. [1998].

Differences between the *PKD1* nucleotide sequence and those of the homologous genes can provide an indication of the age of single nucleotide polymorphisms (SNPs) and the order of duplication events. In the case of M/T1092, the mouse sequence encodes threonine, while both published human *PKD1* sequences, as well as all homologous genes, code for methionine at that amino acid position. The current data would suggest a very recent origin, subsequent to the duplication events, of the back mutation resulting in the human M/T1092 polymorphism.

At position 1649, threonine is conserved between the mouse and human *PKD1* orthologs and in HG2, while the other HGs encode methionine. The findings may indicate that the origin of T/M1649 variant preceded some of the duplication events and that HGs 1, 3, 4, and 5 were derived from the M1649 allele.

The Q/R739 polymorphism is evident in the two published human *PKD1* sequences, with Q739 encoded by sequence #L33242, and R739 encoded by #P98161. The homologous genes show differences. Based on the mouse *PKD1*

sequence (Accession # O08852), one could argue that Q739 is the ancestral allele, which gave rise to HG3 and HG4, whereas HGs 1,2, and 5 appear to be the products of duplication events involving the derived R739 allele.

Reliable conclusions on PKD1 gene genealogy and the origins and age of the HGs have to await the accumulation of SNP frequency data in humans, as well as information on other mammalian species. Such data will also allow conclusions on the role of the homologous genes as reservoirs of mutations in gene conversion events.

### REFERENCES

- Afzal AR, Florentie RN, Taylor R, Patton MA, Saggat-Malik A, Jeffery S. 2000. Novel mutations in the duplicated region of the polycystic kidney disease 1 (PKD1) gene provides supporting evidence for gene conversion. *Genet Test* 4:365-370.
- Aguiari G, Savelli S, Garbo M, Bozza A, Augello G, Penolazzi L, De Paoli Vitali E, La Torre C, Cappelli G, Piva R, del Senno L. 2000. Novel splicing and missense mutations in autosomal dominant polycystic kidney disease 1 (PKD1) gene: expression of mutated genes. *Hum Mutat* 16:444-445.
- American PKD1 Consortium. 1995. Analysis of the genomic sequence for the autosomal dominant polycystic kidney disease gene predicts the presence of a leucine-rich repeat. *Hum Mol Genet* 4:575-582.
- Bear JC, Parfrey PS, Morgan JM, Martin CJ, Cramer BC. 1992. Autosomal dominant polycystic kidney disease: new information for genetic counselling. *Am J Med Genet* 43:548-553.
- Bogdanova N, McCluskey M, Sickmann K, Markoff A, Todorov V, Dimitrakov D, Schiavello T, Thomas M, Kalaydjieva L, Dworniczak B, Horst J. 2000. Screening the 3' region of the polycystic kidney disease 1 (PKD1) gene in 41 Bulgarian and Australian kindreds reveals a prevalence of protein truncating mutations. *Hum Mutat* 16:166-174.
- Bogdanova N, Markoff A, Gerke V, McCluskey M, Horst J, Dworniczak B. 2001. Homologues to the first gene for autosomal dominant polycystic kidney disease (ADPKD) are pseudogenes. *Genomics* 74:333-341.
- Choukroun G, Itakura Y, Man NK, Christophe JL, Albouze G, Jungers P, Grunfeld JP. 1995. The rate of progression of renal failure in ADPKD. *Contrib Nephrol* 115:28-32.
- Cotton RG, Sriver CR. 1998. Proof of "disease causing" mutation. *Hum Mutat* 12:1-3.
- Dalgaard OZ. 1957. Bilateral polycystic kidney disease of the kidneys: a follow-up of two hundred and eighty-four patients and their families. *Acta Med Scand* 328:1-251.
- European Polycystic Kidney Disease Consortium. 1994. The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. *Cell* 77:881-894.
- Gabow PA. 1993. Autosomal dominant polycystic kidney disease. *N Engl J Med* 329:332-342.
- Hughes J, Ward CJ, Peral B, Aspinwall R, Clark K, San Millan JL, Gamble V, Harris PC. 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* 81:289-98.
- Kim UK, Jin DK, Ahn C, Shin JH, Lee KB, Kim SH, Chae JJ, Hwang DY, Lee JG, Namkoong Y, Lee CC. 2000. Novel mutations of the PKD1 gene in Korean patients with autosomal dominant polycystic kidney disease. *Mutat Res* 432:39-45.
- Koptides M, Hadjimichael C, Koupepidou P, Pierides A, Constantinou Deltas C. 1999. Germinal and somatic mutations in the PKD2 gene of renal cysts in autosomal dominant polycystic kidney disease. *Hum Mol Genet* 8:509-513.
- Koptides M, Mean R, Demetriou K, Constantinides R, Pierides A, Harris PC, Deltas CC. 2000. Screening of the PKD1 duplicated region reveals multiple single nucleotide polymorphisms and a de novo mutation in Hellenic polycystic kidney disease families. *Hum Mutat* 16:176.
- Krawczak M, Cooper DN. 1997. The human gene mutation database. *Trends Genet* 13:121-122.
- Loftus BJ, Kim UJ, Sneddon VR, Kalush F, Brandon R, Fuhrmann J, Mason T, Crosby ML, Barnstead M, Cronin L, Deslattes Mays A, Cao Y, Xu RX, Kang HL, Mitchell S, Eichler EE, Harris PC, Venter JC, Adams MD. 1999. Genome duplications and other features in 12 Mb of DNA sequence from human chromosome 16p and 16q. *Genomics* 60:295-308.
- Miller SA, Dykes DD, Polesky HF. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucl Acids Res* 16:12-15.
- Milutinovic J, Fialkow PJ, Agodoa LY, Phillips LA, Rudd TG, Bryant JL. 1984. Autosomal dominant polycystic kidney disease: symptoms and clinical findings. *Q J Med* 53:511-522.
- Parfrey PS, Bear JC, Morgan J, Cramer BC, McManamon PJ, Gault MH, Churchill DN, Singh M, Hewitt R, Somlo S, Reenders ST. 1990. The diagnosis and prognosis of autosomal dominant polycystic kidney disease. *N Engl J Med* 323:1085-1090.
- Peral B, Gamble V, Strong C, Ong AC, Sloane-Stanley T, Zervas K, Winearls CG, Harris PC. 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.
- Perrichot R, Mercier B, Carre A, Cledes J, Feret C. 2000a. Identification of 3 novel mutations (Y4236X, Q3820X, 11745+2 ins3) in autosomal dominant polycystic kidney disease 1 gene (PKD1). *Hum Mutat* 6:582.

- Perrichot R, Mercier B, Quere I, Carre A, Simon P, Whebe B, Cledes J, Ferec C. 2000b. Novel mutations in the duplicated region of PKD1 gene. *Eur J Hum Genet* 8:353-359.
- Perrichot RA, Mercier B, de Parscau L, Simon PM, Cledes J, Ferec C. 2001. Inheritance of a stable mutation in a family with early-onset disease. *Nephron* 87:340-345.
- Peters DJ, Sandkuijl LA. 1992. Genetic heterogeneity of polycystic kidney disease in Europe. *Contrib Nephrol* 97:128-139.
- Phakdeekitcharoen B, Watnick TJ, Ahn C, Whang DY, Burkhart B, Germino GG. 2000. Thirteen novel mutations of the replicated region of PKD1 in an Asian population. *Kidney Int* 58:1400-1412.
- Phakdeekitcharoen B, Watnick TJ, Germino GG. 2001. Mutation analysis of the entire replicated portion of pkd1 using genomic DNA samples. *J Am Soc Nephrol* 12:955-963.
- 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.
- Roscoe JM, Brissenden JE, Williams EA, Chery AL, Silverman M. 1993. Autosomal dominant polycystic kidney disease in Toronto. *Kidney Int* 44:1101-1108.
- Rossetti S, Strmecki L, Gamble V, Burton S, Sneddon V, Peral B, Roy S, Bakkaloglu A, Komel R, Winearls CG, Harris PC. 2001. Mutation analysis of the entire PKD1 gene: genetic and diagnostic implications. *Am J Hum Genet* 68:46-63.
- Rost B, Sander C. 1994. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* 20:55-72.
- Schievink WI, Torres VE. 1997. Spinal meningeal diverticula in autosomal dominant polycystic kidney disease. *Lancet* 349:1223-1224.
- 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.
- Watnick TJ, Phakdeekitcharoen B, Johnson A, Gandolph M, Wang M, Briefel G, Klinger KW, Kimberling W, Gabow P, Germino GG. 1999. Mutation detection of PKD1 identifies a novel mutation common to three families with aneurysms and/or very-early-onset disease. *Am J Hum Genet* 65:1561-1571.
- Watnick TJ, Piontek KB, Cordal TM, Weber H, Gandolph MA, Qian F, Lens XM, Neumann HP, Germino GG. 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, Germino GG. 1998. Somatic mutation in individual liver cysts supports a two-hit model of cystogenesis in autosomal dominant polycystic kidney disease. *Mol Cell* 2:247-251.