GENETIC AND PHYSICAL MAPPING OF AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY DISEASE

SUSAN ELIZABETH POUND

Thesis submitted for the degree of

Doctor of Medicine

University of Edinburgh

December 1993

DEDICATION

I dedicate this thesis to my husband, Dr. James H. Bryce, who sustained me during its production by his love, support and encouragement

DECLARATION

I declare that this thesis was composed entirely by myself, and that the work presented is my own unless otherwise stated.

The following publications are derived from the work in this thesis:-

1. POUND SE, Carothers AD, Pignatelli PM, Macnicol AM, Watson ML, and Wright AF. Linkage disequilibrium between D16S94 and the adult onset polycystic kidney disease (PKD1) gene. (1992). *J. Med. Genet.* 29: 247-248.
2. Pignatelli PM, POUND SE, Carothers AD, Macnicol AM, Allan PL, Watson ML, and Wright AF. Multipoint mapping of adult onset polycystic kidney disease (PKD1) on chromosome 16. (1992). *J. Med. Genet.* 29: 638-641.
3. Wright AF, Teague PW, POUND SE, Pignatelli PM, Macnicol AM, Carothers AD, DeMey RJ, Allan PL, and Watson ML. A study of genetic linkage heterogeneity in 35 adult onset polycystic kidney disease families. (1993). *Hum. Genet.* 90: 569-571.

ACKNOWLEDGEMENTS

My thanks are due to a great many people, many of whom I am unable to mention here but to whom I am grateful nonetheless.

First and foremost my supervisor, Dr. Alan Wright, who has guided and directed me through the maze of molecular genetics, and who has read this thesis on more than one occasion. I very much enjoyed my time spent in Alan's lab. at the MRC Human Genetics Unit, and will always remember the company provided by Lee, Kate, Micheala, Doug, Fiona, Marlene and, especially, John Brown who provided invaluable help and with whom the Independent crossword was mastered daily.

Many thanks are due to Andrew Carothers and Peter Teague for their help with the statistics and linkage programs, to the Unit photography department for producing the photographs, to Anne Macnicol for cups of coffee and clinical information, and to Dr. Michael Watson who suggested the idea in the first place!

Finally, as ever, thanks are due to my parents who have always supported and encouraged me in whatever I do.

This research was funded by a Clinical Training Fellowship from the Medical Research Council.

ABSTRACT

Autosomal dominant polycystic kidney disease (ADPKD) is one of the commonest genetic diseases in man, affecting 1 in 1000 of the general population. It is characterised by the progressive formation and enlargement of multiple cysts in the kidney and other organs leading to deterioration in renal function and the development of hypertension in middle life. End-stage renal disease is a common outcome, leading to a need for dialysis or transplantation. In order to make a significant impact on the understanding and treatment of ADPKD, it is necessary to isolate the gene and determine its product and function.

Linkage was first demonstrated between the gene for ADPKD (PKD1) and the D16S85 locus in chromosomal region 16p13 in 1985, and this has since been amply confirmed. Initial studies failed to show evidence of genetic heterogeneity, but reports have subsequently appeared describing families that fail to show linkage to the PKD1 region. A sample of 35 families with ADPKD from central Scotland, previously typed with two markers from the PKD1 region [3'HVR (D16S85), and CMM65 (D16S84)], were typed with three other markers, VK5 (D16S94), 26-6 (D16S125), and SM7 (D16S283). A multipoint analysis was carried out in order to define the location of PKD1, and to look for evidence of genetic heterogeneity. The results provided statistically significant evidence for the presence of a second locus for ADPKD (PKD2). Two-point and multipoint lod scores were calculated for families shown to be linked to chromosome 16p (designated PKD1 families). Two-point linkage analysis gave highly significant lod scores for each of the five chromosome 16 markers used, and multipoint analysis showed that the most likely position of the PKD1 gene was between CMM65 (D16S84) and SM7 (D16S283). Haplotypes were drawn up for each family, and evidence for linkage disequilibrium between PKD1 and VK5 (D16S94) was found. Eleven individuals with interesting recombinants were identified, and these were analysed in detail using a further set of 8 markers in collaboration with other workers. Nine of these occur between the distally located 3'HVR (D16S85) and CMM65 (D16S84). However, there is one recombinant with CMM65 (D16S84), and one with 26-6 (D16S125), which localises PKD1 to between these markers.

In order to obtain cloned DNA from this genetically defined region of interest, spanning approximately 750 kb of DNA, yeast artificial chromosomes (YACs) were isolated from available libraries. Two human YAC libraries and a mouse YAC library were screened using polymerase chain reaction (PCR) amplification, followed by probing, of pooled YAC DNA. Screening was performed using PCR primers to the GGG1 (D16S259), 26-6 (D16S125), SM33p22, VK5 (D16S94), and ATPL (ATPase-like vacuolar proton channel) loci. In this way, three YACs have been identified: a 280 kb YAC containing the GGG1 locus, a 230 kb YAC containing both the 26-6 and VK5 loci, and a 160 kb mouse YAC containing the ATPL locus.

The three YACs were characterised initially by pulsed field gel electrophoresis (PFGE) purification, followed by digestion with a restriction endonuclease and ligation to catch-linkers in order to amplify YAC fragments by PCR. This amplified material was used to obtain a chromosomal localisation for the YACs using fluorescence *in-situ* hybridisation. The GGG1 YAC was found to be chimaeric and so was not characterised further. PFGE restriction maps were produced for the 26-6/VK5 and ATPL YACs and attempts were made to isolate the end clones. Clusters of CpG rich sites in the 26-6/VK5 YAC suggest the presence of at least two expressed sequences in the cloned DNA.

Experiments were carried out in order to isolate expressed sequences from the human 26-6/VK5 YAC using a YAC-cDNA hybrid selection technique. cDNA libraries were PCR amplified using vector primers, the YAC was amplified using biotinylated primers, and the two annealed at 65^{o}C. YAC-cDNA hybrids were selected by binding to streptavidin-coated magnetic beads, and taken through a further round of selection. Enriched cDNA fractions were isolated and probed back to the YAC and to a variety of control DNA samples. Several of the cDNA fractions appeared to contain repetitive DNA sequences which made them difficult to characterise. Further characterisation of these fractions is required in order to definitively identify expressed sequences present in the YAC, which would therefore be candidate genes for ADPKD.

ABBREVIATIONS

ACE	- angiotensin converting enzyme
ADPKD	- autosomal dominant polycystic kidney disease
ARPKD	- autosomal recessive polycystic kidney disease
ATPL	- ATPase-like vacuolar protein
bp	- base pairs
cAMP	- cyclic AMP
cDNA	- complementary DNA
cen	- centromere
CEPH	- Centre d'Etude du Polymorphisme Humain
DNA	- deoxyribonucleic acid
dNTP	- deoxynucleotide triphosphate
EDTA	- ethylenediamine tetra-acetic acid
EGF	- epidermal growth factor
GFR	- glomerular filtration rate
IPTG	- isopropylthio-β-D-galactoside
kb	- kilobase pairs (10^3 bp)
LMP	- low melting point
Mb	- Megabase pairs (10^6 bp)
mRNA	- messenger RNA
p	- short arm of the chromosome
PCR	- polymerase chain reaction
PFGE	- pulsed field gel electrophoresis
PGP	- phosphoglycolate phosphatase
PKD	- polycystic kidney disease
PKD1	- the ADPKD gene located on chromosome 16p
	- ADPKD families demonstrating linkage to chromosome 16p
PKD2	- the ADPKD gene(s) located elsewhere in the genome
	- ADPKD families which are unlinked to chromosome 16p
q	- long arm of the chromosome
RFLP	- restriction fragment length polymorphism

RNA	- ribonucleic acid
TBE	- Tris borate EDTA
tel	- telomere
tRNA	- transfer RNA
θ	- recombination fraction
X-gal	- 5-bromo-4-chloro-3-indolyl-β-D-galactoside
YAC	- yeast artificial chromosome
Z	- lod score

TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

1.1 INTRODUCTION

End-stage renal disease, leading either to long term renal replacement therapy with dialysis or to renal transplantation, is a major financial burden on the health service and an important cause of morbidity in the general population. In Europe, the most common causes of end-stage renal disease are glomerulonephritis and diabetes mellitus, closely followed by adult-onset polycystic kidney disease and hypertension. This thesis is concerned with autosomal dominant polycystic kidney disease, and the localisation of the gene responsible for the majority of cases of this disease.

Adult-onset (autosomal dominant) polycystic kidney disease is one of the most common genetic diseases in man, and is an important cause of chronic renal failure. Since the biochemical defect responsible is unknown, "positional cloning" techniques are being employed in an attempt to isolate the gene(s) involved. Positional cloning describes the isolation of a disease gene on the basis of its chromosomal location. Cytogenetic abnormalities in gene carriers are extremely useful in determining the chromosomal location of a gene, especially when patients with overlapping deletions are available. However, polycystic kidney disease has not been associated with visible cytogenetic abnormalities. Under these circumstances, linkage analysis is usually the first step in chromosomal localisation of a disease gene. Genetic markers selected from throughout the genome are tested for co-segregation with the disease, and highly polymorphic probes are used to maximise the efficiency of the search. Once a gene has been localised to a particular chromosomal region, high resolution genetic mapping can define the boundaries of this region so that physical mapping techniques such as pulsed field gel electrophoresis and cosmid and yeast artificial chromosome contig mapping can then be employed. In order to clone the gene, various strategies are used such as

searching for CpG islands, isolating cDNAs from the area of interest, and identifying expressed sequences that are conserved across species. DNA from affected patients and from controls is then screened to detect any mutations present. Disease-causing mutations may be detected at more than one locus (genetic heterogeneity) especially in patients from different populations.

If the processes involved in the pathophysiology of a disease have been elucidated, a candidate gene approach can be taken in which a known gene, whose function suggests that it may be involved in the disease process, is screened for mutations in affected patients. This has so far not been possible in the case of polycystic kidney disease. Where possible, a combination of linkage analysis followed by a candidate gene approach may be most effective. This has been demonstrated, for example, in the case of autosomal dominant retinitis pigmentosa (ADRP), where some families show genetic linkage of ADRP to chromosome 3q, the site of the rhodopsin gene, which led directly to the identification of rhodopsin mutations (Dryja *et al.*, 1990).

1.2 RENAL CYSTIC DISEASES

There are many causes of renal cysts, both inherited and acquired (Zerres *et al.*, 1984). The inherited causes are summarised in Table 1.1 and, of these, by far the most common is autosomal dominant polycystic kidney disease (ADPKD) which will be fully discussed in later sections. Other causes are autosomal recessive polycystic kidney disease (ARPKD), medullary cystic kidney disease, and renal cystic disease occurring as part of a recognised syndrome (e.g. trisomy 13, tuberous sclerosis, von Hippel Lindau disease, and the rare autosomal recessive Meckel's and Jeune's syndromes). A full list can be found in Zerres *et al.* (1984).

In ARPKD there are two phenotypes, which are distinguished by the age at presentation and whether renal or hepatic manifestations predominate. Renal disease predominates in patients presenting in the neonatal period or during infancy. The most severely affected infants present with Potter's syndrome (oligohydramnios, secondary pulmonary hypoplasia, congenital contractures and

Table 1.1 INHERITED RENAL CYSTIC DISEASES

Disease	Inheritance	Frequency	Clinical features	Carrier detection
Adult polycystic kidney disease	Autosomal dominant	1 in 1000	Renal cysts, hypertension, berry aneurysms, chronic renal failure.	Ultrasound scan, genetic linkage studies
Infantile polycystic kidney disease	Autosomal recessive	1 in 10,000 to 1 in 40,000	Renal and hepatic cysts, chronic renal failure, hepatic fibrosis	Not available
Tuberous sclerosis	Autosomal dominant	1 in 10,000	Skin lesions, epilepsy, mental retardation, renal cysts, cardiac rhabdomyosarcomas.	Skull X-ray, Wood's light CT scan of brain and viscera
von Hippel Lindau disease	Autosomal dominant	Rare	Retinal angioma with cerebellar angioblastoma. Hypernephroma. Cysts in kidney, liver and pancreas.	CT scans of brain and abdominal viscera. Retinal examination.
Medullary cystic disease	Autosomal dominant	Rare	Adult onset. Uraemia, anaemia, renal salt wasting, medullary cysts.	Ultrasound, renal function tests.
Juvenile nephronophthisis	Autosomal recessive	1 in 50,000	Anaemia, renal salt wasting,uraemia medullary cysts, retinal degeneration, growth retardation.	Not available
Orofaciodigital syndrome type I (OFD I)	X-linked dominant	Rare	Oral frenulae and clefts, hamartomata of the tongue, digital malformation, often polycystic kidneys.	Not available

Table 1.1 summarises some of the most common causes of inherited renal cystic disease. A fuller list can be found in Zerres *et al.* (1984).

flattened facies), and die of respiratory failure soon after birth. Patients whose disease develops in childhood or adolescence have predominantly hepatic problems (e.g. hepatic cysts and fibrosis leading to portal hypertension and hypersplenism) and less obvious kidney manifestations. Both types can usually be easily distinguished from patients with ADPKD by their pattern of inheritance and the high degree of liver involvement, although in very early onset ADPKD the presence of renal cysts prenatally, or in infancy, can cause diagnostic confusion. There is considerable overlap between the two phenotypes of ARPKD, and there may be very variable expression within one family. As yet, the gene for ARPKD has not been mapped.

There are various causes of acquired renal cystic disease, two of which are common and may potentially cause diagnostic difficulty in cases of ADPKD. Renal cysts may occur at any age in the presence of uraemia. They develop as chronic renal failure progresses, and are particularly common in dialysis-dependent patients. In a study by Dunnill *et al.* (1977), 14 out of 30 patients on long term intermittent maintenance haemodialysis had bilateral cystic disease, the main complications of which were haemorrhage and tumour formation. In these cases the kidneys are usually shrunken, rather than enlarged as in ADPKD. Simple renal cysts are common in later years and are found in healthy individuals. They do not lead to progressive renal failure and, as discussed later, rarely lead to problems in diagnosing ADPKD.

The best example of an environmental influence on cyst formation is seen in the CFWwd mouse (Werder *et al.*, 1984), discussed further in section 1.5.1. Approximately 70% of these animals develop renal cysts, but this does not occur if they are kept in a germ free environment. Renal cystic disease may also result from chronic exposure to chemicals and drugs that are known to be cystogenic. These include antioxidants such as diphenylamine, alloxan, streptozotocin, lithium chloride and cis-platinum. All of these agents can produce bilateral uniform renal cysts in otherwise normal kidneys (Gardner, 1988). Endogenous conditions, such as obstruction to flow with resultant back pressure, can result in the development

of renal cystic disease. Examples of this include the development of a unilateral congenital multicystic-dysplastic kidney, which is a non-heritable form of renal dysplasia, and the development of cystic kidneys following dialysis or transplantation (Gardner, 1988).

1.3 AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY DISEASE (ADPKD)

This is the most common form of hereditary renal disease. There is progressive expansion and an increasing number of renal cysts, which can arise from any part of the nephron, leading to bilateral renal enlargement and, in a large proportion of cases, irreversible renal failure. Much of the early information regarding polycystic kidney disease came from a detailed study of the Danish population by Dalgaard (1957). He found the lifetime risk of adult-onset polycystic kidney disease (PKD) to be 1 per 1000 with marked variability in phenotype, although there was often a striking similarity within families. The incidence of adult-onset PKD may be higher than 1 in 1000, however, since autopsy studies have shown that there are likely to be as many clinically silent cases as there are cases leading to symptomatic renal impairment (Hatfield and Pfister, 1972). In Dalgaard's study of 284 patients (Dalgaard, 1957), the mean age of onset of the disease was found to be 40.7 years and the mean age of death 51.5 years, with the main cause of death being uraemia. The inheritance of ADPKD is autosomal dominant with complete penetrance by the age of 80 years. Dalgaard (1957) calculated the rate of new mutations to be 6.5×10^{-5} to 12×10^{-5} per gene per generation, which was of the same order of magnitude as the highest rates found for other genetic diseases. However, prior to the development of ultrasound scanning, relatively poor means of pre-symptomatic diagnosis were available, and therefore some parents thought to be "unaffected" may have been gene carriers. Using age at death as a marker for disease severity, no significant evidence was found to support the presence of anticipation (i.e. an increasing severity of phenotype in successive generations). More recently, the gene responsible for the majority of cases of ADPKD has been localised to the short arm of chromosome 16 (Reeders *et al.*, 1985), and genetic locus heterogeneity has been

demonstrated (Romeo *et al.*, 1988; Kimberling *et al.*, 1988). Approximately 86% of families with ADPKD demonstrate linkage with 16p markers (PKD1 families) (Peters and Sandkuijl, 1992), the remainder being unlinked to 16p (PKD2 families).

1.3.1 Pathology of ADPKD

The metanephros (permanent kidney) is derived from both mesoderm and ectoderm. The nephron originates from the nephrogenic cord or metanephric blastema (mesoderm). The collecting ducts, calyces, pelvis and ureters are derived from the ureteric bud (ectoderm). During the 4th to 5th week of gestation the ureteric bud arises from the mesonephric duct, and at about the 6th week, branches to form the primitive pelvis and major calyces of the kidney. The nephrons are formed from the ureteric bud and the metanephric blastema at 8 weeks. The blastema gives rise to nephrogenic cells, which result in glomeruli and tubules, and stromatogenic cells which become the interstitium. Nephrogenesis stops at birth and no new nephrons can be produced later in life. In contrast, stromatogenesis continues to occur in adult life.

The ureter enters the kidney at the hilum, dilates into the pelvis, and then divides into the major and minor calyces. On cut surface, the kidney is made up of a cortex and medulla. The medulla consists of renal pyramids, the apices of which are called papillae, each related to a calyx. The functioning units of the kidney are the nephrons (Figure 1.1). They consist of a glomerulus and a renal tubule, and span the cortex and medulla. The glomerulus is a capillary network which is invested in Bowman's capsule; the invaginated blind proximal end of the renal tubule that has a visceral and a parietal layer. The glomerular filter is made up of three layers; a thin layer of fenestrated capillary epithelial cells, the glomerular basement membrane and the visceral epithelial cells of Bowman's capsule (podocytes). The renal tubule consists of a proximal convoluted tubule, loop of Henle, distal convoluted tubule and collecting duct. The collecting duct runs to the tip of the renal papillae and opens into the calyces.

Figure 1.1

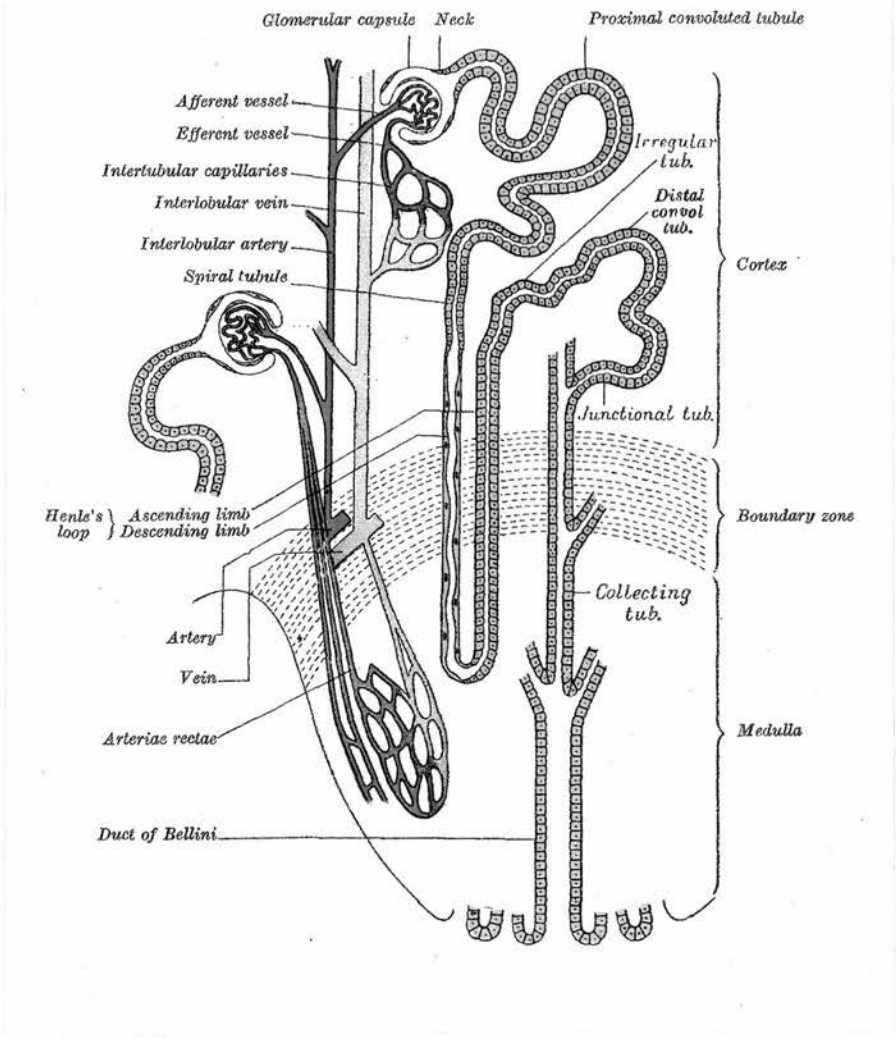


Figure 1.1 demonstrates the anatomy of a normal nephron, which consists of a glomerulus, proximal convoluted tubule, loop of Henle, distal convoluted tubule and collecting tubule. The nephron spans the cortex and medulla and empties into the calyceal system. The blood supply to the nephron is also demonstrated.

The structure of normal renal tubular epithelial cells varies considerably at different levels of the nephron. Proximal tubular cells are highly developed cuboidal cells with long microvilli, numerous mitochondria and extensive intercellular interdigitations. A sodium (Na^+) pump is located in the basal membrane. The descending limb and the proximal part of the ascending limb of the loop of Henle are lined by flat epithelium, and the distal part of the ascending limb is lined by cuboidal cells. The distal tubule is lined by cuboidal cells that have fewer mitochondria than those of the proximal tubule and do not possess brush borders.

In ADPKD there is diffuse involvement of the kidneys by cysts that are usually spherical and unilocular, ranging from a few millimetres to several centimetres in size. The kidneys usually retain a reniform shape, and the cysts are distributed throughout the cortical and medullary regions. The pyramids are difficult to identify, and the calyces and pelvis often greatly distorted. Most of the cysts are filled with a clear fluid, but they may also be filled with haemorrhagic or purulent material. The percentage of the nephrons undergoing cystic change is estimated to be less than 1% (Birenboim *et al.*, 1987; Grantham *et al.*, 1987).

On light microscopy, normal renal tissue is found among the cysts although it often shows the effects of compression by adjacent cysts. There may also be secondary glomerular sclerosis, tubular atrophy and interstitial fibrosis. The cyst epithelium is usually thin and undifferentiated in appearance. Focal hyperplasia of the epithelial lining of the cysts has recently been shown to be a common finding (Grantham *et al.*, 1987), and this may be such as to cause formation of polyps protruding into the cyst lumen. Electron microscopy shows that the cyst-lining epithelial cells are usually devoid of apical microvilli, contain a small number of organelles and lie on a basement membrane that varies inconsistently from normal to thickened, or extensively laminated. Two types of epithelial lining cell can be distinguished, depending on the composition of the cyst fluid. 'Proximal' epithelial lining cells have smooth surfaces, indistinct cell margins, squamoid or low cuboidal form and non-dilated intercellular channels. 'Distal' cyst epithelium has a

cobblestone surface, distinct cell margins, taller cuboidal cells and prominently dilated intercellular channels.

1.3.2 Radiology of ADPKD

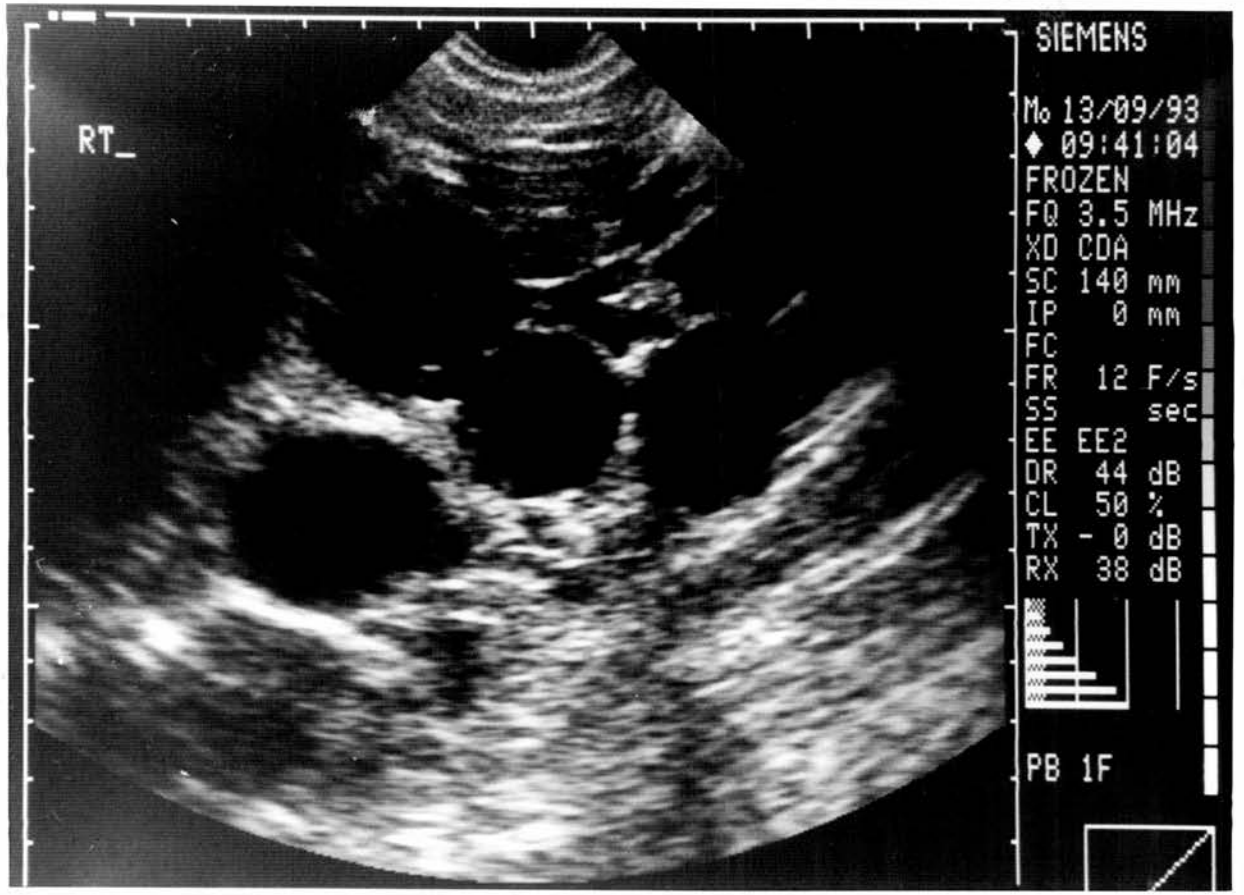
The diagnosis of ADPKD depends on the finding of bilateral renal cysts. Ultrasound scanning is now widely used as the most convenient and sensitive method of screening. The ultrasonic appearance of a polycystic kidney is shown in Figure 1.2. Since ADPKD shows age-dependent penetrance, the interpretation of negative findings on ultrasound is also age-dependent, and genetic linkage analyses and risk calculations require knowledge of the sensitivity and specificity of negative screening results at a given age. Figures obtained by Bear *et al.* (1992) estimate the probability of false negative diagnosis in PKD1 families to be 36% in the first decade, 8% in the second and third decades and 0% thereafter. Other studies have shown that a positive ultrasound diagnosis always correlates with the genotype, and that all gene carriers more than 30 years old have ultrasound-detectable renal cysts (Parfrey *et al.*, 1990; Coto *et al.*, 1992).

One of the main differential diagnoses is that of multiple simple renal cysts. Simple cysts are multiple in 50% of cases and bilateral in 26%. They also show an increasing frequency with age, being present in at least 27% of people over the age of 50 years. Fortunately, the presence of simple cysts is exceptional under the age of 30, and uncommon under the age of 40 (Tada *et al.*, 1983), so that they rarely present a problem in the diagnosis of ADPKD.

1.3.3 Clinical features of ADPKD

ADPKD is predominantly a renal disease, but there are many non-renal manifestations. The penetrance of ADPKD is 100% by the age of 80 years (Dalgaard, 1957). Phenotypic heterogeneity in ADPKD is marked both within and between families: variation exists in the age of onset, rate of loss of glomerular filtration rate (GFR), degree of hypertension, prevalence of extrarenal cysts and cardiac valve

Figure 1.2



This figure shows a polycystic kidney from an ADPKD patient as seen on ultrasound scanning. The anterior abdominal wall is at the top of the photograph, and several larger cysts are clearly seen.

abnormalities, and incidence of subarachnoid haemorrhage. The prognosis is also highly variable.

The mean age of onset of symptoms is in the fourth decade, although a minority of patients develop renal failure in the neonatal period. The most common complications are the development of hypertension and renal failure. Overall, 36% of males and 7% of females with ADPKD are asymptomatic and normotensive, but in those under the age of 30 years this is higher, at 66% of males and 11% of females. In women, the most frequent clinical manifestations are urinary tract infection (present in 69%), and hypertension (present in 61%): in men, urinary tract infection is present in 19% and hypertension in 42% (Milutinovic *et al.*, 1984). Other presenting symptoms are flank pain, haematuria, and nephrolithiasis. Hypertension and symptoms such as haematuria and back pain, but not urinary tract infection, correlate well with renal size on ultrasound (Milutinovic *et al.*, 1984). There is no increase in the incidence of renal tumours in ADPKD patients (Dalgaard, 1957). As in adults, cysts are detected in children by ultrasound scanning. Gross haematuria, hypertension, and a palpable abdominal mass are present in more than 90% of children with a positive ultrasound scan (Sedman *et al.*, 1987).

Hypertension is common in patients with ADPKD, and usually predates the onset of renal failure. Parfrey *et al.* (1990) found that hypertension was present in 22% of all ADPKD patients under the age of 20 years, 26% aged 20-39, 55% aged 40-59, and 75% over the age of 60. Hypertensive ADPKD patients manifest more severe cystic involvement than normotensive patients, suggesting a relationship between structural alteration and hypertension prior to the onset of renal failure (Gabow *et al.*, 1987). It has been hypothesised that cystic involvement could lead to local ischaemia and increased formation of renin-angiotensin. Support for this was obtained by showing a significantly increased rise in plasma renin activity, in response to angiotensin converting enzyme (ACE) inhibitors, in hypertensive compared with normotensive ADPKD patients (Bell *et al.*, 1988). Tubulo-cystic epithelium has the potential to synthesise renin, and elevated levels are seen in

both epithelial cells and cyst fluid (Torres *et al.*, 1992). Thus it seems likely that renin is, in some way, involved in the pathogenesis of hypertension.

The most important consequence of ADPKD is renal failure. This is rare under the age of 30, but end-stage renal disease is present in 50-75% of patients by the age of 70 years (Churchill *et al.*, 1984; Parfrey *et al.*, 1990), the mean age of onset being 59 years (Parfrey *et al.*, 1990). As in other forms of chronic renal failure, the rate of decline of renal function is adversely affected by hypertension.

Extrarenal involvement is common in ADPKD. Liver cysts are found in 29% of patients with ADPKD over the age of 10 years (Milutinovic *et al.*, 1980). They are not found in patients at risk for ADPKD who have not yet developed renal cysts. The prevalence of liver cysts increases with age and with decreasing GFR. They are generally asymptomatic but can cause pain if they are very large or become infected. Impaired liver function is rare. Cysts are also found in other organs such as pancreas, spleen and ovary (Dalgaard, 1957).

According to post mortem studies, the prevalence of intracranial aneurysms in patients with ADPKD is 10-30% (Levey *et al.*, 1983). Rupture of such aneurysms is one of the most serious complications of ADPKD, and may occur early in the third decade before hypertension or renal failure have developed (Chauveau *et al.*, 1990). Patients from families with a history of subarachnoid haemorrhage are at a higher risk of intracranial aneurysm than those without a family history (Kaehny *et al.*, 1987), and patients who have had a ruptured aneurysm are at high risk of a recurrence (Chauveau *et al.*, 1990). Routine screening of asymptomatic patients is not cost effective when the method of screening is cerebral angiography (Levey *et al.*, 1983), and is unlikely to be so even with non-invasive techniques such as magnetic resonance imaging (Wiebers and Torres, 1992). However, screening should be offered to certain subgroups of patients, e.g. those with a family history of intracranial aneurysm or subarachnoid haemorrhage, those undergoing major elective surgery with anticipated haemodynamic instability and those with symptoms suggesting the presence of an aneurysm (Wiebers and Torres, 1992). Up to 20% of patients also have other cardiovascular abnormalities, the most common

of which are structural defects of the aortic and mitral valves, potentially leading to valve regurgitation and heart failure. There is also an increased frequency of congenital cardiac defects, e.g. bicuspid aortic valves (Leier *et al.*, 1984). Histology of the affected valves shows myxomatous degeneration with loss and disruption of collagen fibres. A similar histological picture and an increase in bicuspid and floppy valves are also seen in Ehlers-Danlos syndrome types I and III, in which there are known to be abnormal collagen fibrils.

1.3.4 Prognosis of ADPKD

In a recent study of patients with ADPKD by Gabow *et al.* (1992), 71% were alive at age 50, 53% at age 58, and 23% at age 70. This is in keeping with other studies (Churchill *et al.*, 1984; Parfrey *et al.*, 1990). Children diagnosed in the neonatal period appear to have a poor prognosis, whereas those diagnosed later in childhood seem to have a benign early course (Sedman *et al.*, 1987). The most common cause of death directly attributable to ADPKD is end-stage renal disease, although rupture of an intracranial aneurysm can cause early death in some patients, prior to the onset of renal failure. Gabow *et al.* (1992) found that the following variables were associated with a worse prognosis as regards progression to renal failure: younger age at diagnosis, male gender, hypertension, increased left ventricular mass, hepatic cysts in females, more than three pregnancies, gross haematuria, urinary tract infections in men, renal size, and inheritance of the PKD1 mutation. The following were not associated with a worse renal prognosis: gender of the affected parent, mitral valve prolapse, intracranial aneurysm, less than three pregnancies, hepatic cysts in men, and urinary tract infections in females.

Reproductive fitness in ADPKD appears to be high. Dalgaard (1957) found that 65% of women with ADPKD became pregnant one or more times, and a high percentage of live births resulted. A more recent study has shown that, in patients with ADPKD, reproduction rate, spontaneous abortion, stillbirth and symptoms consistent with a urinary tract infection were no different from those in normal women (Milutinovic *et al.*, 1983).

In general terms, it is difficult to find any clinical distinction between ADPKD linked to chromosome 16p (PKD1) and the form of ADPKD that is unlinked to chromosome 16 (PKD2). The two families in which non-linkage to chromosome 16 was first identified (Romeo *et al.*, 1988; Kimberling *et al.*, 1988) had clinically typical ADPKD. Parfrey *et al.* (1990), however, found the average age of onset of end-stage renal disease to be 57 years in PKD1 patients compared with 69 years in non-PKD1 patients. Renal cysts also tended to develop later in life, and hypertension and renal impairment were less frequent and occurred later. In this study, however, PKD1 was not invariably more severe, and the observations on non-PKD1 patients came from only two, albeit large, families. This could therefore reflect allelic heterogeneity rather than locus heterogeneity. Further evidence for a difference in prognosis between PKD1 and PKD2 comes from Bear *et al.* (1989), who noted that the proportion of individuals at 50% risk of ADPKD who had cysts was lower at every age in PKD2 families than in PKD1 families. Ravine *et al.* (1992) studied 18 PKD1 and 5 non-PKD1 families, and found that non-PKD1 patients lived longer than PKD1 patients, had a lower risk of progressing to renal failure or developing hypertension, were diagnosed at an older age and had fewer renal cysts at diagnosis. In all of these studies an ascertainment bias could be present, since an unknown proportion of PKD1 and PKD2 families may have mild disease which does not require referral to a renal clinic.

1.3.5 Treatment of ADPKD

No specific treatment is available to retard the progression of cyst formation. Effective control of blood pressure provides one of the few means of delaying the progression of renal failure. In view of the postulated involvement of the renin-angiotensin axis, angiotensin converting enzyme (ACE) inhibitors seem a logical choice of drug in the management of hypertension. The fact that Lisinopril, an ACE inhibitor, produces an acute fall in renal vascular resistance without a fall in GFR (Watson *et al.*, 1992) adds weight to the suggestion that ACE inhibitors may be of particular value in the treatment of hypertension associated with ADPKD. In

general, however, hypertension is treated as it would be in essential hypertension, the main aim being good control of blood pressure.

Conservative management of chronic renal failure includes strict attention to blood pressure control, dietary restriction, particularly of protein and phosphate, and treatment of acidosis and anaemia. Sooner or later, end-stage renal disease will occur and patients will then require long term renal replacement therapy (dialysis) or renal transplantation.

If cysts are extremely large, decompression may relieve pain but it does not slow the progression of renal failure. Cysts may become infected, and any infections must be rigorously treated with appropriate antibiotics.

Therefore, although measures such as the control of hypertension and the availability of renal replacement therapy and transplantation have greatly improved the prognosis of ADPKD, the hope is that the localisation of the gene, followed by an understanding of the biochemical defect resulting from mutation within it, will allow the development of treatments to actually retard cyst formation.

1.3.6 Genetic counselling in ADPKD

As in any disease inherited as a fully penetrant autosomal dominant trait, every offspring of an affected parent has a 50% chance of developing the disease. Since patients with ADPKD usually do not develop symptoms until the third or fourth decade of life, the ability to diagnose the disease at an early stage is of great benefit when counselling individuals regarding their prognosis. The mainstay of early diagnosis is ultrasound scanning; the rates of false negative diagnosis being estimated as 36% below the age of 10 years, 8% or less thereafter until the age of 30 years when false negative diagnosis is unlikely (Bear *et al.*, 1992). Linkage of PKD1 to 3'HVR (D16S85) and the development of flanking markers around this gene permit accurate identification of individuals at risk of developing the condition, provided that the kindred is large enough to establish linkage and the disease haplotype. In non-recombinants, the accuracy of diagnosis in PKD1 families was estimated to be 99% using flanking markers (Breuning *et al.*, 1990b). In families

where the disease is unlinked to chromosome 16, no such genetic diagnosis is possible at present.

Early diagnosis has both advantages and disadvantages (Watson *et al.*, 1990). It allows screening for the development of hypertension and, if appropriate, for the presence of intracranial aneurysms. Disadvantages include difficulty in obtaining employment and loading of life assurance premiums. The availability of ante-natal diagnosis raises other ethical problems. Although the majority of patients are in favour of ante-natal diagnosis, few would consider termination of pregnancy on the basis of the result (Macnicol *et al.*, 1991). Parental curiosity is unlikely to be in the best interests of the child and, in view of the potential disadvantages, it is only in rare cases that the diagnosis need be established under the age of 18 years (Watson *et al.*, 1990). Counselling, particularly with a view to educating patients with regard to polycystic kidney disease, is very effective in improving knowledge and understanding of the disease, thereby affecting attitudes towards diagnosis and treatment (Macnicol *et al.*, 1991).

1.4 GENETICS OF ADPKD

1.4.1 Linkage analysis

Genetic linkage is present when a genetic marker is found to segregate non-randomly with a disease or other genetic phenotype in families. It occurs when two loci lie close enough on the same chromosome that they tend to be inherited together. Linked loci may segregate apart as a result of physical exchange of genetic material between paired chromosomes during meiosis. The frequency of such recombination between two loci is a measure of the genetic distance between them. Recombination frequencies can also be used to order loci in relation to each other, and thereby establish their position within a specific chromosomal region.

The lod score statistic (Z) is a measure of the likelihood that linkage is present, and the maximum likelihood value of the lod score (Z_{max}) is conventionally used to indicate the recombination fraction (θ_{max}) between two loci. The lod score is calculated as:-

$$Z(\theta) = \log 10 \frac{\text{likelihood of linkage data assuming a specified recombination fraction } (\theta)}{\text{likelihood assuming the loci are unlinked (i.e. } \theta = 0.5)}$$

Linkage was first detected between ADPKD and the DNA probe 3'HVR, in a study of nine families with the typical clinical features of the disease (Reeders *et al.*, 1985). The 3'HVR (D16S85) locus is a highly polymorphic microsatellite sequence which is located 8 kb downstream of the α-globin locus (Jarman *et al.*, 1986). The α-globin locus had been mapped to the short arm of chromosome 16 (Barton *et al.*, 1982), and is now known to lie between 170 and 450 kb from the telomere (Wilkie *et al.*, 1991). It consists of an array of 17 base-pair tandem repeats, and polymorphism at the locus is attributed to variability in the number (70-450) of these repeats (Jarman *et al.*, 1986). The maximum likelihood lod score for linkage of ADPKD to 3'HVR was 7.31 at a recombination fraction (θ) of 0.07 in males and 16.41 at $\theta = 0.04$ in females: this gave a combined lod score of 25.85 at $\theta = 0.05$ (Reeders *et al.*, 1985), suggesting that 3'HVR and ADPKD are separated by approximately five centiMorgans (5cM). Linkage of ADPKD to the short arm of chromosome 16 (16p) was confirmed by the presence of tight linkage with the phosphoglycolate phosphatase (PGP) locus (Reeders *et al.*, 1986; Watson *et al.*, 1987). No recombinants were found between the disease and PGP in either study. Breuning *et al.* (1987a) mapped human α-globin to 16pter-p13.3 distal to PGP, using an unbalanced translocation which carried one allele at the 3'HVR (D16S85) locus and two alleles at the PGP locus.

The ADPKD disease gene located on chromosome 16 has been designated PKD1. The position of the PKD1 locus relative to 3'HVR (D16S85) was shown in a multipoint linkage analysis (Reeders *et al.*, 1988), which indicated that the odds in favour of PKD1 lying proximal to 3'HVR (D16S85) were greater than 10,000 : 1 compared with a distal location. By a combination of linkage and somatic cell hybrid analyses, this group also showed that the DNA markers CRI-0327 (D16S63) and CRI-090 (D16S45) were located proximal to both 3'HVR (D16S85) and PKD1, this order being 170 times more likely than an order placing PKD1 between CRI-

0327 (D16S63) and CRI-090 (D16S45). The disease gene was found to be flanked by 3'HVR (D16S85) distally and 24-1 (D16S80) proximally with a recombination rate of 8% between the two markers (Breuning *et al.*, 1987b), this order being at least 1000 times more likely than either of the other two possibilities. The detection of flanking markers allowed improved diagnosis by genetic linkage analysis. The positions of markers located on chromosome 16p are shown in Figure 1.3.

Once linkage to a particular region of a chromosome has been established, detailed genetic and physical mapping is required. Using a combination of cell lines containing rearranged chromosomes, cytogenetically-defined breakpoints and multipoint linkage analysis, genetic and physical maps of an array of polymorphic markers flanking PKD1 have been constructed (Reeders *et al.*, 1988; Breuning *et al.*, 1990a; Germino *et al.*, 1990). The genetic localisation of PKD1 was further refined by the identification of recombination events between close flanking markers and PKD1 (Somlo *et al.*, 1992b). Three crossovers were identified that placed PKD1 proximal to GGG1 (D16S259), and two that placed PKD1 distal to 26-6 (D16S125). All these crossovers were in affected individuals from unequivocal PKD1 families. PKD1 was therefore localised to a region between GGG1 (D16S259) and 26-6 (D16S125). Within this span, only 92.6SH1.0 recombines with the disease locus, and three recombinants from different families have been identified. In the same two meioses that place PKD1 distal to 26-6 (D16S125), the data also place PKD1 distal to 92.6SH1.0, while in a single unaffected case (negative ultrasound at age 33 years) the disease locus appears to be proximal to 92.6SH1.0.

1.4.2 Genetic heterogeneity of ADPKD

A reliable diagnosis based on linkage analysis assumes that the disease is caused by mutation(s) at only one locus. The original localisation of ADPKD to chromosome 16 was made by demonstrating linkage to 3'HVR (D16S85) and PGP in only nine families (Reeders *et al.*, 1985 and 1986). Other early studies confirming PKD1 linkage were also carried out on similarly small numbers of families (Watson *et al.*,

Figure 1.3 ORDER OF MARKER LOCI ON CHROMOSOME 16p

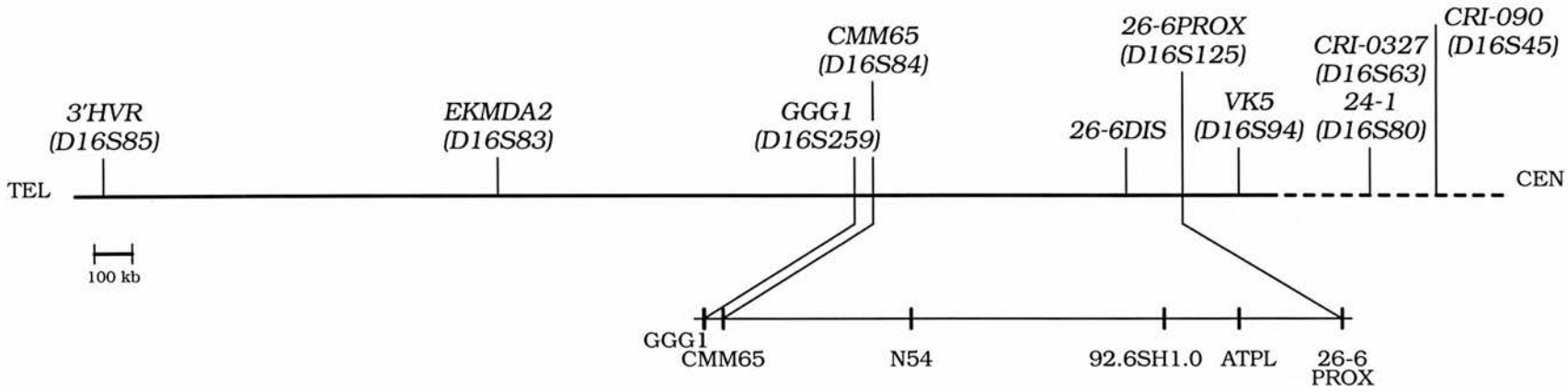


Figure 1.3 is a map of markers on the short arm of chromosome 16 and their relationship to the PKD1 locus. The map is taken from Somlo *et al.* (1992) and Breuning *et al.* (1990a). Markers between GGG1 and 26-6PROX have been ordered on cosmid contig clones. The positions of the markers are to scale except for those with the dashed line.

1987; Lazarou *et al.*, 1987). None of these studies found any evidence to suggest genetic heterogeneity of ADPKD.

The presence of genetic locus heterogeneity would lead to problems in ante-natal and pre-symptomatic diagnosis, genetic counselling, and in the molecular characterisation of the disease gene itself. It is therefore very important to establish whether or not heterogeneity is present and, if so, what proportion of families is unlinked to chromosome 16. In order to answer this question, Reeders *et al.* (1987) extended their original study to analyse 28 northern European pedigrees, and again found no evidence to suggest the presence of a second locus for ADPKD. Following this, analysis of two large families of Italian origin raised the possibility of heterogeneity (Romeo *et al.*, 1988; Kimberling *et al.*, 1988). Both families had polycystic kidney disease with clinical features indistinguishable from those seen in families linked to chromosome 16. Linkage analysis, however, provided no evidence of linkage to 3'HVR (D16S85) in either pedigree. These original reports have been followed by other case reports detailing single kindreds with ADPKD which show no linkage to chromosome 16 (Brissenden *et al.*, 1989; Bachner *et al.*, 1990; Nørby and Schwartz, 1990). Further studies of linkage heterogeneity in the Italian population (Mandich *et al.*, 1990; Turco *et al.*, 1991) have failed to show any evidence of heterogeneity, suggesting that the proportion of unlinked families is small. Studies of seven Icelandic families with ADPKD revealed that four were PKD1 families showing linkage to chromosome 16p, and that three were non-PKD1 families unlinked to chromosome 16 (Fossdal *et al.*, 1993).

The above reports of unlinked families argue convincingly for the presence of a second locus for ADPKD, although two authors have argued that "recombinational heterogeneity", i.e. an unusual clustering of recombination in a minority of families, should be considered as an alternative explanation (Hulten, 1988; Elles *et al.*, 1990). The more families that are found to be unlinked to chromosome 16, the less likely is this explanation, and the more likely that there is indeed at least one other locus responsible for the ADPKD phenotype.

Studies with larger numbers of families in different populations are required to establish the proportion of ADPKD families that are unlinked to chromosome 16. A study of 42 families by Pieke *et al.* (1989) obtained an estimate of 96% of families linked to chromosome 16 (with a lower 95% confidence limit of 93%). Peters and Sandkuijl (1992) analysed data from 328 ADPKD families (available as a result of the European Concerted Action on polycystic kidney disease and including data presented in this thesis) and obtained a value of 86% linked to 16p, although the 95% confidence limits were wide (79-91%).

The presence of genetic heterogeneity means that caution must be exercised in providing ante-natal and pre-symptomatic diagnosis. Flanking markers are of value in separating linked from unlinked families by discriminating between single and double recombinants. In small families where there is insufficient data to establish linkage with α-globin, genetic prediction will have an error roughly equal to the proportion of unlinked families.

1.4.3 Genetic Localisation of PKD2

Until very recently, little progress had been made in defining the chromosomal location of the gene or genes responsible for ADPKD in families where there is no demonstrable linkage to chromosome 16. Nørby and Schwartz (1990) have reported a possible locus on chromosome 2. In a large Danish family with ADPKD unlinked to α-globin, a positive but not statistically significant lod score of 2.12 was obtained between the disease and YNH24 (D2S44), a marker on the short arm of chromosome 2. A large Sicilian ADPKD kindred of over 250 members, shown to be unlinked to chromosome 16, was analysed using 29 marker loci on 11 autosomes (Kumar *et al.*, 1991). In this family, the data exclude linkage of the disease locus to 17 marker loci, including YNH24 reported above, and show no evidence of close linkage with the other loci. Peters *et al.* (1993) have now reported evidence for linkage between PKD2 and the chromosome 4 markers D4S423 and D4S231. The map locations of these markers are 4q21-4q23 (Weissenbach *et al.*, 1992) and 4q21-4q24 (Mills *et al.*, 1992) respectively. Two-point analysis of eight families gave a

cumulative lod score of 20.11 at 8% recombination for D4S423 and of 16.04 at 5% recombination for D4S231. The family reported by Nørby and Schwartz (1990) as showing linkage to YNH24 (D2S44), provides stronger evidence of linkage to chromosome 4 with a maximum lod score of 4.26 (Peters *et al.*, 1993).

PKD2 has therefore been localised to chromosome 4q, although mutations at more than two loci may yet be responsible for the ADPKD phenotype.

1.4.4 Physical mapping of PKD1

As mentioned above, once detailed genetic maps of the region had been established, physical mapping techniques were used to verify the position of markers in relation to each other, and to determine the physical distances between them. The probe CMM65 (D16S84) was found to show no genetic recombination with PKD1, and thus somatic cell hybrids containing spontaneous chromosomal breakpoints were used to construct a physical map of this region (Germino *et al.*, 1990). This group localised CMM65 (D16S84) to between CRI-0327 (D16S63) proximally and α-globin distally with odds of >1000 : 1, EKMDA2 (D16S83) to between CMM65 (D16S84) and 3'HVR (D16S85) also with odds of >1000 : 1, and PKD1 to between EKMDA2 (D16S83) and CRI-0327 (D16S63) with odds of >500 : 1, the most likely location being at the CMM65 (D16S84) locus. Harris *et al.* (1990) constructed a library from DNA enriched for a 320 kb *Not*I fragment containing CMM65 (D16S84). New probes from this region were isolated and used together with existing markers to construct a long range restriction map that encompassed the distal markers for PKD1, including α-globin. DNA surrounding the α-globin complex is GC-rich, and this GC-rich region was found to extend for more than 2 Megabases proximal to α-globin (Harris *et al.*, 1990). Since unmethylated CpG islands have been correlated with the presence of genes (Lindsay and Bird, 1987), it is probable that this region will be rich in transcribed sequences.

As discussed above, Somlo *et al.* (1992b) localised PKD1 to between GGG1 (D16S259) distally and 26-6PROX (D16S125) proximally, after identifying recombinants with these loci. A long range restriction map linking these two

markers has been constructed (Germino *et al.*, 1992). A *Not*I linking library for this region was made from a radiation hybrid containing fragments of human chromosome 16, and the clones mapped onto a panel of somatic cell hybrids (Himmelbauer *et al.*, 1991). Bi-directional cosmid walking was then initiated from three DNA markers (CMM65, 26-6, and VK5) and four of the *Not*I linking clones (Germino *et al.*, 1992). Three contigs of 500 kb, 150 kb and 60 kb have been obtained, separated by gaps of less than 20 kb and approximately 60 kb. The maximum distance between the flanking markers GGG1 (D16S259) and 26-6PROX (D16S125) was found to be less than 750 kb (Germino *et al.*, 1992). Two probes, originally thought to recognise single-copy loci, were discovered to hybridise to more than one locus on chromosome 16 (Germino *et al.*, 1992). 26-6 (D16S125) recognises two loci, 26-6PROX and 26-6DIS, less than 150 kb apart. 26-6PROX is the polymorphic locus (Gillespie *et al.*, 1990). There is also a 40 kb segment, beginning approximately 10 kb telomeric to the *Not*I clone N54, that is present in several copies on chromosome 16 (Germino *et al.*, 1992). Cosmids representing several of these repetitive loci were found in chromosome 16 enriched libraries. The copy within the PKD1 region can be distinguished from all others by virtue of a different sized *Eco*RI fragment that hybridises to a *Bam*HI subclone of the repetitive segment (pBB1.9).

Restriction analysis of genomic and cosmid DNA identified a high density of rare cleavage sites within the PKD1 region (Harris *et al.*, 1990; Germino *et al.*, 1992), many of which represent unmethylated CpG islands. CpG islands usually coincide with the 5' end of genes (Lindsay and Bird, 1987), and can therefore be used as markers to speed up the search for genes. However, since not all genes are associated with CpG islands, this approach runs the risk of missing some genes. The large number of CpG clusters mapped to the interval suggests that the PKD1 region will be gene-rich, and already at least 20 sets of non-overlapping cDNA clones have been isolated from the region (Germino *et al.*, 1992).

1.4.5 Human - mouse homology

Groups of genes tend to remain physically associated in different species during evolution, and many of these conserved linkage groups have been identified between human and mouse. The presence of a conserved linkage group may benefit the search for a human gene in several ways. A correlation may be established between a known mouse mutation and a human hereditary disease with a similar phenotype. For example, as discussed below, the *cpk* mouse is a good model for human ARPKD, and the mutation causing polycystic kidneys in that strain has been mapped to mouse chromosome 12 (Davisson *et al.*, 1991). This would suggest that the homologous human ARPKD gene could be localised to either human chromosome 2p23-p25 or 7q22-q31. Alternatively, homologous candidate genes may be identified if a mouse mutant with a similar phenotype has been attributed to a particular gene.

The gene for human ADPKD that is linked to the α-globin locus on chromosome 16 (Reeders *et al.*, 1985) is reported to be located in a 750 kb region flanked by GGG1 (D16S259) and 26-6 (D16S125) (Germino *et al.*, 1992). Himmelbauer *et al.* (1992), took two *Not*I linking clones and two human cDNA clones from this 750 kb region, along with the two flanking markers, and found that all of them mapped to mouse chromosome 17. There is, therefore, an area of linkage conservation between human chromosome 16p and mouse chromosome 17, and it is likely that the mouse homologue of PKD1, if it exists, will be included in the conserved region. It is of interest that the expressed α-globin gene, on the distal side of PKD1 in the human, is located on chromosome 11 in the mouse, whereas Hba-ps4, an α-globin pseudogene, is located near to the conserved region on mouse chromosome 17, but has no equivalent in the human genome.

1.5 ANIMAL MODELS OF POLYCYSTIC KIDNEY DISEASE

1.5.1 Murine models

There are no naturally occurring mouse models with a similar phenotype to ADPKD, but the following models are available. Table 1.2 summarises the

Table 1.2 CYSTIC DISEASES OF THE KIDNEY IN MAN AND MOUSE

Disease/Mutation	Mode of inheritance	Species	Human chromosome	Mouse chromosome
ADPKD (PKD1)	Dominant	Man	16p13.3	17
ADPKD (non-PKD1)	Dominant	Man	4	?
cpk/cpk	Recessive	Mouse	2 or 7	12
pcy/pcy	Recessive	Mouse	3	9
CBA/N	? X-linked	Mouse	? X	? X
SR2-3 (chimaeric)	Dominant	Mouse	?	11
SV40 T-antigen (transgenic)	Dominant	Mouse	?	?
SBM (transgenic)	Dominant	Mouse	?	?

Table 1.2 shows the chromosomal location, where known, of the genes involved in human ADPKD and various mouse models. These are further described in the text. The table is taken from Reeders (1992), Mackay *et al.* (1987) and Rahilly *et al.* (1992).

chromosomal location of the genes involved in human ADPKD and various mouse models.

CFWwd mouse - this strain of mouse develops a form of ADPKD, with progressive development of cystic lesions in Bowman's capsule, proximal and distal collecting tubules (Werder *et al.*, 1984). The cystic disease is inherited as an autosomal dominant trait with reduced penetrance. There is clear environmental modulation of cyst formation and enlargement since only 4% of animals raised in a germ-free environment develop PKD compared with 70% in a standard environment.

C57BL/6J-*cpk*/*cpk* (*cpk*) mouse - this strain transmits a lethal form of PKD in an autosomal recessive fashion. The *cpk* mouse model has been the one most intensively studied since its clinical and gross morphological characteristics closely resemble those of human ARPKD, although abnormalities are limited to the kidney. The *cpk* mutation has been mapped to mouse chromosome 12 (Davisson *et al.*, 1991).

DBA/2-*pcy*/*pcy* (*pcy*) mouse - this strain develops slowly progressive PKD similar to human ADPKD, but transmitted in an autosomal recessive manner (Takahashi *et al.*, 1986). Renal cysts develop in all segments of the nephron, and progressively enlarge with age. No extrarenal cysts are seen.

CBA/N mouse - this mouse strain has an immunodeficiency phenotype mapped to a recessive mutation (*xid*) on the X chromosome. Renal cystic lesions occur which closely resemble those seen in human ADPKD: all animals greater than 3 months old show cystic dilatation affecting all segments of the nephron. The renal cystic lesions also appear to be X-linked (Rahilly *et al.*, 1992).

SV40 T-antigen (T-ag) model - mice transgenic for the complete SV40 early region encoding the small and large T-antigen display a phenotype characterised by

choroid plexus tumours leading to death by 3 or 4 months of age. They also display renal abnormalities including focal glomerulosclerosis and tubular cysts, with renal expression of the SV40 large T-ag (MacKay *et al.*, 1987).

SBM transgenic mouse - SBM mice were produced by microinjection into the mouse genome of a hybrid construct containing the SV40 enhancer, β globin promoter, and murine c-*myc* proto-oncogene (Trudel *et al.*, 1991). They consistently develop PKD and die of renal failure. The phenotype resembles human ADPKD in autosomal dominant transmission, complete penetrance, bilateral renal cysts occurring throughout the nephron, epithelial hyperplasia and progressive renal insufficiency. Differences include the absence of extrarenal manifestations of ADPKD and the presence of haemopoietic malignancies in some mice.

Chimaeric mouse - the introduction of SR2-3 embryonic stem cells into mouse embryos leads to the production of chimaeric mice, which reproducibly exhibit the characteristics of ADPKD, i.e. the formation of fluid-filled cysts in the kidney transmitted in a dominant manner (Boulter *et al.*, 1992). The cysts are lined with epithelial cells and occur in both medulla and cortex. SR2-3 cells have a single integrated copy of the v-*src* oncogene within the SR2 retroviral vector. Again, the model differs from human PKD in that no cysts are seen in other organs, despite the fact that these tissues were found to be chimaeric. The integrated oncogene may itself be responsible for the disorder, and there are parallels with transgenic mice over-expressing the c-*myc* proto-oncogene. Alternatively, the disease may result from a dominant mutation at the ADPKD locus, either arising spontaneously in the SR2-3 cell line or caused by the insertion of the retroviral vector. The site of insertion of the transgene has now been mapped to mouse chromosome 11.

1.5.2 Rat models

A colony of Sprague-Dawley rats (Han: SPRD) has been bred which inherits polycystic kidney disease as an autosomal dominant trait. It has several features resembling human ADPKD including chronic progressive uraemia, renal tubular basement membrane abnormalities, development of cysts in all nephron segments, and evidence of abnormal cell proliferation with increased expression of the c-*myc* proto-oncogene in cystic epithelium (Cowley *et al.*, 1993). It is possible to breed two sub-strains of Han: SPRD, one that exhibits polycystic degeneration of the kidneys and a rapid rise in serum urea, serum creatinine and blood pressure, and one that exhibits the typical histological features of ADPKD but no loss of renal function or rise in blood pressure. Since they have a larger circulating blood volume than mice, they are particularly useful in animal studies of progression of renal failure (Gretz *et al.*, 1992).

1.5.3 Feline model

A model for ADPKD is available in a Persian cat (Biller and Pflueger, 1990) which closely resembles human ADPKD in its phenotype and genetic transmission.

1.5.4 Cell culture models

Human monolayer ADPKD epithelia - human ADPKD cells grow easily in culture (Wilson *et al.*, 1986). Although the epithelial cells contain few organelles, they remain polarised with apical tight junctions and microvilli facing upwards in the dish. They have been validated as a model system for the analysis of cellular alterations in PKD.

MDCK (Madin-Darby canine kidney) cell line in collagen gel - this is a model for cystogenesis using dog kidney cells. Three-dimensional clonal growth in a hydrated collagen gel yields spherical monolayered cysts. They exhibit epithelial cell proliferation, intracavity fluid accumulation, and matrix remodelling (Grantham *et al.*, 1989). Only about 7% of the cultured MDCK cells participate in cyst formation.

1.6 PATHOPHYSIOLOGY OF ADPKD

The pathophysiology of cyst formation is not understood, but several studies looking at different aspects of cyst biology have provided insights into the kind of processes likely to be involved (reviewed by Gabow, 1991). A renal tubule cyst is an enlarged segment of the nephron consisting of a fluid-filled sac lined by a single layer of epithelium. It seems that cysts can form in response to genetic, non-genetic, and toxic influences, and therefore it is likely that several pathogenetic pathways converge to produce a single phenotype. Cysts are separated by apparently normal areas of renal parenchyma and it seems that renal failure develops because the enlargement of established cysts compromises the function of adjacent tubules, rather than because more tubules become involved in the cystic process (Grantham *et al.*, 1987). Microdissection studies of ADPKD kidneys in the early stages of cystic change show that cysts can originate from any segment of the nephron, and are randomly distributed with normal nephrons between them (Baert, 1978). Aspiration of renal cysts from ADPKD patients reveals that some cysts contain fluid with concentrations of sodium and creatinine approximately equal to those of plasma, whereas others have very low concentrations of sodium but increased concentrations of creatinine (Gardner, 1969). These patterns resemble those found in fluid from normal proximal and distal renal tubules respectively.

Electron microscopic studies show that a relatively small proportion of cysts remain attached to the tubule segments, and thus transepithelial net solute and fluid secretion into the cavity must occur (Grantham *et al.*, 1987). The same study showed that mean cell surface area did not increase in direct proportion to cyst diameter, and thus epithelial hyperplasia must be a central element in the progressive enlargement of cysts. In some cysts this hyperplasia was accentuated by projections into cyst lumina of polyps, and cord-like arrangements of cells.

Cellular proliferation without secretion would lead to solid tumour formation, whereas increased cellular proliferation accompanied by increased cellular secretion, which in turn is accompanied by increased intracyst pressure (Grantham *et al.*, 1989), would produce outward expansion yielding a cyst.

Therefore, it seems that the two basic criteria necessary for cyst formation are increased epithelial cell proliferation and altered fluid transport. Epithelial proliferation can be the result of alteration of extracellular matrix composition; altered susceptibility to, or synthesis of, peptide growth factors; or abnormal expression of growth-related genes including proto-oncogenes. Fluid accumulation in renal cysts could be the result of alterations in ion pumping, or reflect changes in polarity of cell structure or function.

1.6.1 Epithelial cell hyperplasia

As discussed above, epithelial cells proliferate during cyst formation (Grantham *et al.*, 1987). Primary cultures of ADPKD cystic epithelial cells show an increased cell growth potential when compared with normal cells (Wilson *et al.*, 1986). These cells are also able to attach and proliferate without extracellular matrix, a prerequisite for normal renal tubular cell growth (Wilson *et al.*, 1992). Epithelial cell hyperplasia is also seen in the kidneys of SBM transgenic mice (Trudel *et al.*, 1991).

1.6.2 Growth factors in ADPKD

Epidermal growth factor (EGF) is a potent mitogen, acting via a high affinity receptor that possesses intrinsic tyrosine kinase activity. Wilson and Sherwood (1991) demonstrated an exaggerated proliferative response to EGF in cultured ADPKD cystic epithelium, and also found EGF to be present in cyst fluid and cystic epithelial cells. They also showed that transforming growth factor β (TGFβ), an inhibitor of renal tubular proliferation, did not produce the expected inhibition in cultured ADPKD cystic epithelium.

1.6.3 Oncogenes in ADPKD

Elevated proto-oncogene expression is common to a number of different models of polycystic kidney disease, such as the c-*myc* transgenic (SBM) mouse, SV40 large T-antigen transgenic mouse, *cpk* mouse and Han: SPRD rat. SBM kidneys contain markedly elevated levels of c-*myc* mRNA in comparison to other transgenic organs

and normal mouse kidney controls. While the SV40 enhancer sequence contains an element known to cause expression in kidney cells, the entire SV40 enhancer has been shown to be active in many cell types *in vitro*, and in a variety of tissues in transgenic mice. Trudel *et al.* (1991) postulated that the novel coupling of the β-globin promoter sequences and the SV40 enhancer in the SBM construct modifies the tissue specificity of the enhancer. The restriction of cystic changes to the kidney therefore parallels the apparent renal-limitation of transgene expression. The identification of revertant mice with mutations in the transgene provides evidence that the PKD phenotype is a direct result of the SBM transgene, with overexpression of c-*myc* in the renal tubular epithelium and consequent abnormal cell proliferation. This strongly suggests that cyst formation can arise through the dysregulation of tubular epithelial cell proliferation.

In the Han: SPRD rat strain, there is increased epithelial cell proliferation overlying areas of basement membrane thickening. *In situ* hybridisation demonstrates high levels of c-*myc* mRNA in cyst epithelia (Cowley *et al.*, 1993). In addition, the *cpk* mouse strain, which models ARPKD, shows elevated expression of the c-*myc* oncogene with a 2-6 fold increase in c-*myc* mRNA at 2 weeks, and a 25-30 fold increase in c-*myc* mRNA at 3 weeks of age (Cowley *et al.*, 1987). Since the cystic kidneys show only a minimal increase in cell proliferation over normal kidneys, the level of c-*myc* expression is out of proportion to the level of cell proliferation, suggesting that elevated c-*myc* expression may be involved in the pathogenesis of PKD. Oncogene expression was also induced to a lesser degree by other stimuli to renal cell growth, such as folic acid treatment and unilateral nephrectomy with consequent compensatory hypertrophy of the other kidney.

The SR2-3 cell line has integrated into its genome a retroviral vector carrying the v-*src* oncogene. Chimaeric mice derived from this cell line develop polycystic kidneys, but the role of v-*src* expression is unclear, since the levels of v-*src* activity in the kidneys are low. Moreover, a chimaeric mouse derived from a second ES cell line carrying the same retroviral vector had a similarly low level of v-*src* activity in the kidneys but no evidence of cystic kidney disease (Boulter *et al.*,

1992). This would suggest that the v-*src* oncogene does not have a direct role in cyst formation in this model.

1.6.4 Fluid transport defects in ADPKD

A series of experiments by Mangoo-Karim *et al.* (1989) focused on the role of cyclic AMP (cAMP) as a chemical mediator of secretion. In the absence of cAMP, the MDCK cells did not form cysts. If a cAMP phosphodiesterase inhibitor was added, allowing low levels of cAMP to accumulate, cell proliferation was initiated, although this only resulted in balls of cells, not cysts. Substances such as prostaglandin E_1 (PGE_1) and cholera toxin, which markedly increase intracellular cAMP, did however result in large fluid-filled cysts. It therefore seems that although low levels of cAMP were able to cause proliferation, it was the fluid secretion generated by high levels of cAMP which led to cyst formation.

In the above study, it was also noted that ouabain completely inhibited fluid secretion. Human cystic ADPKD epithelial tissue appears to demonstrate altered cellular polarity with Na^+-K^+-ATPase located exclusively on the apical rather than the basolateral cell surface (Wilson *et al.*, 1991). Actin and the ankyrin-fodrin complex, which are believed to be involved in the anchoring of Na^+-K^+-ATPase in the appropriate membrane, co-localised with Na^+-K^+-ATPase at the apical surface: a variety of other basolateral and apical proteins maintained their appropriate positions. Avner *et al.* (1992) studied cyst formation in the *cpk* mouse, and found that in control and cystic proximal tubules Na^+-K^+-ATPase distribution was restricted to the basolateral membrane of cells. However, during early formation of both normal and cystic collecting tubules, there was apical membrane distribution which was transient in the normal cells but persistent in the cystic epithelial cells. Thus, apical expression of Na^+-K^+-ATPase seems to be a normal transient feature of early collecting tubule development, and its persistence in cystic epithelial cells may be a manifestation of the relatively undifferentiated phenotype of epithelial cells lining cysts. Na^+-K^+-ATPase maintains its function in this altered location,

and there is thus an altered direction of cellular sodium transport from the basolateral to the apical cell surface (Wilson *et al.*, 1991).

1.6.5 Extracellular matrix abnormalities in ADPKD

Wilson *et al.*, (1986), demonstrated that ADPKD cells in culture produced greater amounts of extracellular matrix than normal human kidney epithelia. At least one component of the increased matrix appeared to be a proteoglycan. The alteration in extracellular matrix could contribute to cystogenesis via alterations in cell-matrix interactions, such as adhesion, growth and differentiation. Wilson and Sherwood (1991) have shown that type IV collagen mRNA levels increase late in the course of ADPKD. Actin is also excessively abundant in ADPKD renal tissue (Wilson and Sherwood, 1991), as well as being mislocated. Rocco *et al.* (1992) studied levels of transcripts in *cpk* mice, and found that those for growth factors, structural proteins, and fibronectin and laminin receptors were normal. However, there was an early reduction of mRNA levels in the kidney for E-cadherin (uvomodulin) and N-CAM (N-cell adhesion molecule). The presence of E-cadherin and N-CAM appears to guide the sequential differentiation and polarisation of normal renal epithelia, and their attenuated expression in the kidneys of *cpk* mice may be a factor in the pathogenesis of cyst formation.

Thus, many possible mechanisms have been implicated in cystogenesis, but little information is available concerning the sequence of events resulting in the phenotypic expression of renal tubular cysts. The postulated involvement of the extracellular matrix and cytoskeleton in the pathogenesis of ADPKD is particularly attractive in that it may help to explain the extrarenal manifestations of ADPKD. Once the abnormal gene and its function have been identified, this must be related to the abnormal cellular events in order to determine which are primary and which are secondary; and to determine how they relate to the non-renal manifestations of the disease.

The question remains as to why less than 5% of nephrons undergo cystic change (Grantham *et al.*, 1987) despite, presumably, the expression of the gene in all cells. One possibility, is that ADPKD could be an example of a "two-hit" model of pathogenesis (Reeders, 1992). The model for this mechanism is retinoblastoma (Knudson, 1971). This suggests that, in addition to the inherited predisposition, a second sporadic event is required to initiate cyst formation; the sporadic event , for example, being a somatic mutation in the chromosome that does not carry the inherited mutation (Reeders, 1992). Another possibility is that the PKD1 gene contains an heritable unstable repeat element such as that seen in fragile X syndrome, myotonic dystrophy and Huntington's Disease. In this case, repeat number instability at the somatic level may explain the sporadic distribution of cysts (Reeders, 1992). One of the features of the above disorders, however, is anticipation which classically does not occur in ADPKD.

1.7 ISOLATION OF CANDIDATE GENES

Once a disease gene has been localised to a particular chromosomal region, isolation of the gene requires cloned DNA. Standard DNA technology makes use of bacteriophage lambda (λ) and cosmid cloning vectors. Bacteriophage λ is a complex virus of *Escherichia coli* (*E. coli*) which will accept up to a maximum of 23 kb of foreign DNA. Cosmid vectors have been constructed from plasmids containing a fragment of λ DNA which includes the *cos* site, and they are able to accept foreign DNA inserts of 32-47 kb. Use of these vectors to clone a gene-containing region of DNA therefore requires that this region is first reduced to a manageable size, preferably a few hundred kilobases and certainly less than 1 Megabase (1000 kb) of DNA. In 1987, Burke *et al.* described yeast artificial chromosomes (YACs) capable of cloning several hundred to over one thousand kilobase pairs of DNA in the host *Saccharomyces cerevisiae.* This theoretically allows large regions of the human genome to be cloned much more quickly and easily. They can also be used for the stable cloning of DNA fragments that have proven impossible to isolate from cosmid libraries (Coulson *et al.*, 1988). Several YAC libraries have been constructed, both

total genomic and chromosome-specific, and some of these libraries are available for screening. Once one or more cosmids or YACs have been obtained, they can be used to screen tissue-specific cDNA libraries in a variety of ways in order to detect transcribed sequences. These transcribed sequences are candidate genes. Mutation screening is then performed in order to either prove or disprove an association with a disease phenotype.

1.7.1 Candidate genes for PKD1

More than 20 non-overlapping cDNAs have been identified, and these appear to encode parts of at least 18 transcripts, most of which are associated with CpG islands (Somlo *et al.*, 1992a). Northern analysis suggests that most of the genes in the PKD1 region are expressed in many tissues, including the kidney. Housekeeping genes are generally associated with CpG islands, so that identifying genes on the basis of proximity to islands may bias against the identification of tissue-specific genes. Many tissue-specific genes also have islands, however, although the proportion that do so is still unknown. Therefore, in selecting genes for further study from amongst such a large number, suggested functions based on sequence analysis become important. Sequence analysis of most of the cDNAs in the PKD1 region has not suggested likely functions for most of them (Somlo *et al.*, 1992a). However, one of the genes (AJ1) associated with a CpG island, identified by a combination of cosmid walking and chromosome jumping (Gillespie *et al.*, 1990), shows 77% sequence homology with the gene encoding the 16-kiloDalton proteolipid subunit of bovine chromaffin granule H^+-ATPase (Gillespie *et al.*, 1991). In a region where there are major differences between the human and bovine sequences, there is complete identity between the human, *Drosophila*, *Torpedo*, and yeast sequences. In view of possible mechanisms of cyst formation (discussed above), AJ1 seemed an ideal candidate gene. However, the coding region was not mutated when compared with allelic cDNAs isolated from a primary culture of cyst epithelium obtained from an ADPKD patient, and Northern blot analysis suggests that there are no differences in expression of this gene in normal compared with

ADPKD kidneys. These data, and the finding of two recombinants with 92.6SH1.0 placing PKD1 distal to this marker (Somlo *et al.*, 1992b), were felt to exclude AJ1 as the PKD1 gene. It should be noted, however, that DNA from only one patient was screened for the presence of a mutation, and it is conceivable that this patient may be from a PKD2 family especially as the linkage data is not shown. The evidence for the exclusion of AJ1 as a candidate gene for PKD1 is therefore inconclusive.

Using the series of cosmids isolated by Germino *et al.* (1992), the same group searched for deletions, insertions or other rearrangements by screening digested genomic DNA with a minimally overlapping set of cosmids between GGG1 (D16S259) and 26-6 (D16S125) and looking for identifiable restriction fragment length differences. Between 10 and 30 mutated chromosomes from unrelated affected individuals were studied, and no rearrangements detected, although three small segments were not effectively screened (Somlo *et al.*, 1992a). It is therefore likely that techniques capable of detecting single base alterations will be required to detect PKD1 mutations.

1.8 AIMS OF THE PROJECT

This project is concerned with the gene for ADPKD which has been localised to the short arm of chromosome 16 (PKD1). The aims of the project fall broadly into two categories:-

1. Extension of genetic mapping studies in central Scottish ADPKD families, with two-point and multipoint analyses to (i) confirm localisation of the PKD1 gene, (ii) search for evidence of genetic heterogeneity in this population, (iii) identify and assess new PKD1 recombinants, and (iv) search for evidence of linkage disequilibrium.
2. Isolation of yeast artificial chromosomes by screening of the available YAC libraries with a series of probes from the region of 16p known to flank PKD1. Characterisation of the YACs obtained would be followed by screening of cDNA libraries in order to identify candidate genes for PKD1.

CHAPTER TWO

MATERIALS AND METHODS

2.1 LINKAGE ANALYSIS

2.1.1 Ascertainment and diagnosis of family members

Families were ascertained through the Medical Renal Unit, Royal Infirmary of Edinburgh and the Renal Unit, Western Infirmary, Glasgow. Diagnostic criteria included a family history of ADPKD (more than one affected member) and ultrasound findings of two or more cysts greater than 0.5 cm in diameter in one kidney and at least one such cyst in the contralateral kidney (Watson *et al.*, 1987). Pedigrees were drawn up after interviews with patients and were checked against the centralised Register for Births, Marriages, and Deaths for Scotland. Ethical approval was secured and, after obtaining informed consent, 30-50 ml EDTA-anticoagulated blood was drawn and frozen at -70°C prior to DNA extraction. Thirty-five probands and 343 family members were initially ascertained, and their relatives sampled for biochemical and DNA analysis and examined clinically as described previously (Watson *et al.*, 1987). Ultrasound examinations were then arranged and carried out by experienced sonologists, one in each centre. Positive scans were graded according to the following criteria:-

Grade 1 1-5 cysts bilaterally

Grade 2 6-10 cysts bilaterally

Grade 3 11-15 cysts bilaterally

Grade 4 > 16 cysts bilaterally.

2.1.2 Extraction of DNA from blood

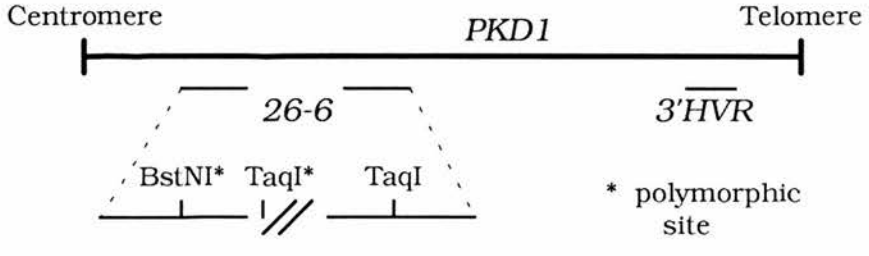
DNA extraction from EDTA-anticoagulated frozen blood was carried out by Mrs. M. Deans according to the method of Kunkel *et al.* (1982). DNA samples were stored at -70°C in TE buffer (10 mM Tris-HCl, 1mM EDTA pH 8).

2.1.3 Markers used in linkage analysis

Probes 3'HVR (D16S85), CMM65 (D16S84), and VK5 (D16S94) were used to type all family members as described by Pignatelli *et al.* (1992). 3'HVR is a hypervariable locus located in 16p13.3 which detects multiple alleles with several restriction enzymes (Jarman *et al.*, 1986). DNA samples were digested with *Pvu*II and allele frequencies were each set at 0.1 in the linkage programs. CMM65 (D16S84) is localised to 16p13.3 and, on probing DNA digested with *Pvu*II, detects two alleles of size 3.5 kb (allele 1) and 1.3 kb (allele 2), with frequencies of 0.37 and 0.63 (Nakamura *et al.*, 1988). VK5 (D16S94), also localised to 16p13.3, detects alleles of 1.6 kb (allele 1) and 1.3 kb (allele 2) with the enzyme *Msp*I, with frequencies of 0.55 and 0.45 (Hyland *et al.*, 1990). The majority of these typings were done by Ms. P. Pignatelli.

26-6 is a single copy 432 bp *Pst*I subclone in pKUN which detects a *Taq*I RFLP on 16p13.3 (Breuning *et al.*, 1990c). Alleles E1 and E2 have frequencies of 0.27 and 0.73 respectively (Breuning *et al.*, 1990a). Approximately 10 kb of this locus is duplicated, with about 120 kb between these duplicated segments (Gillespie *et al.*, 1990). PCR primers have been designed which allow amplification of both loci (Saris *et al.*, 1990), and rapid detection of a RFLP (E1, E2) in the proximal locus by the enzyme *BstN*I (Figure 2.1).

Figure 2.1



The 26-6 PCR primers amplify both the proximal (26-6PROX; D16S125) and distal (26-6DIS) loci. The polymorphic *BstN*I site is present in the proximal locus some 200 bp from the polymorphic *Taq*I site, which is outside the amplified fragment (M. H. Breuning, unpublished results). The distal locus contains a non-polymorphic *Taq*I site.

The families were typed for 26-6 (D16S125) using PCR amplification and digestion of the PCR product with *TaqI* and *BstNI*, as described below.

SM7 (D16S283) is a polymorphic microsatellite (Harris *et al.*, 1991), and families were typed using PCR amplification of the CA repeat followed by separation of the alleles on a 6% polyacrylamide sequencing gel. There are eleven alleles of the following sizes with their frequencies shown in brackets: 107 bp (0.005), 101 bp (0.015), 99 bp (0.06), 97 bp (0.045), 95 bp (0.055), 93 bp (0.04), 91 bp (0.565), 89 bp (0.14), 87 bp (0.025), 85 bp (0.045) and 81 bp (0.005) (Harris *et al.*, 1991). In typing the families, the reference allele (0) was taken to be the 91 bp allele. In order to allow the multipoint linkage analysis to run using the LINKMAP program, the SM7 (D16S283) results were entered into the pedigree files using a maximum of five alleles. Alleles typed as -2, 0, +2, +4, or +6 were left as such, and the less common alleles re-designated as one of the above so that full informativeness was maintained in each family.

2.1.4 Programs used for linkage analysis

Linkage results were analysed using the LINKAGE program package version 5.03 (Lathrop *et al.*, 1984). The ADPKD gene frequency was set at 0.0005, and male and female mutation rates assumed to be equal at 5×10^{-5} per locus per gamete. The female : male recombination ratio was assumed to be constant and was found to maximise the likelihood at a ratio of 0.2. This value was used for the LINKMAP runs. MLINK was unable to support a fixed ratio of recombination in the two sexes, so sex averaged values were obtained. The probabilities of ultrasonographic detection in gene carriers were assumed to be 0.64, 0.92, and 1.00 for ages 0-10 years, 10-30 years and over 30 years respectively (Bear *et al.*, 1992).

Two-point linkage analyses were obtained using the MLINK subroutine. Multipoint analyses were run using LINKMAP with a fixed order of marker loci as follows: D16S85-0.06-D16S84-0.01-D16S283-0.005-D16S125-0.005-D16S94. The Turbo-Pascal source code was modified and recompiled as required. Heterogeneity analyses were carried out by P. Teague using the HOMOG program (Ott, 1985).

2.2 POLYMERASE CHAIN REACTION

Unless otherwise stated, PCR was carried out using 1 unit of Promega *Taq* DNA polymerase in 1 x Promega PCR buffer (50 mM KCl, 10 mM Tris-HCl pH 9.0, 1.5 mM $MgCl_2$, 0.1% Triton X-100), with 200 µM each dNTP (Pharmacia, Ultrapure), and 1.0 µM each primer in a total volume of 25 µl or 50 µl. Samples were overlaid with mineral oil (Sigma) and cycled in either a Cetus or a Hybaid DNA thermal cycler.

2.2.1 Extraction of PCR primers

The majority of primers were made within the Unit (Medical Research Council, Human Genetics Unit, Edinburgh), and were extracted from the deprotected ammonium hydroxide mixture using the following protocol. To 350 µl oligonucleotide mix were added 0.1 volumes 3M sodium acetate pH 6.0, followed by 2 volumes absolute ethanol. The oligonucleotides were allowed to precipitate at -20°C for 30 minutes, and collected by centrifugation at 12,000 rpm for 15 minutes at 4°C. The supernatant was carefully removed and the pellet washed twice using 80% ethanol. Traces of ethanol were removed using a vacuum dessicator, and the pellet resuspended in 200 µl TE buffer. The absorbance of light at 260 nm (A_{260}) was measured, and the concentration of primer calculated according to the following formula:-

$$\text{Concentration} = \frac{A260/ml}{[nA \times 16{,}000]+[nG \times 12{,}000]+[nC \times 7{,}000]+[nT \times 9{,}600]}$$

where n = the number of bases (A, G, C and T) present in the oligonucleotide.

Biotinylated primers were obtained from Oswel DNA Services Ltd., University of Edinburgh, and were HPLC purified and delivered ready for use.

2.2.2 PCR amplification of 26-6 (D16S125)

The duplicated 26-6 locus was amplified from 500 ng genomic DNA using the primer sequences below:-

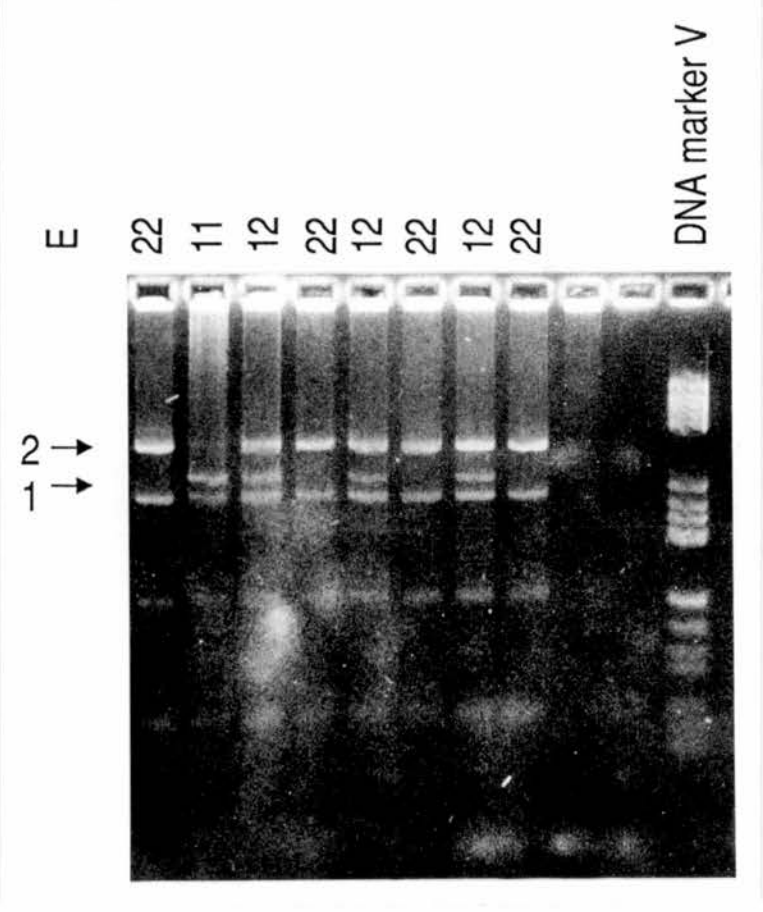
R1: 5' CTG CAG AAT AAA CAA GTG TGT GGT TTC 3'

M4: 5' GGG AAA AAA CAG GTT AAA GAA TGG GAC 3'

The total volume was 25 µl, and $MgCl_2$ concentration 2.5 mM. PCR was carried out for 30 cycles with 1 min at 94°C, 1 min at 65°C, and 2 min at 72°C.

Once the PCR reaction had terminated, the mineral oil was removed, 3 µl 10 x buffer 'B' (Boehringer Mannheim) added, and the amplified DNA digested at 65°C for 1 hour with 5 units *Taq*I (Boehringer Mannheim). 3.5 µl 10 x NEB buffer 2 was then added, and the DNA digested for 2 hours at 37°C with 5 units *BstN*I (New England Biolabs). The products were then separated on a 3% Nusieve (FMC):1% LMP agarose gel (Figure 2.2).

Figure 2.2 *BstN*I polymorphism of the 26-6PROX (D16S125) locus



*BstN*I polymorphism of unrelated individuals, showing the E1 and E2 alleles of 26-6 (D16S125). The DNA marker V (Boehringer Mannheim) is *Hae*III-digested pBR322

2.2.3 PCR amplification of SM7 (D16S283)

The SM7 microsatellite CA repeat was amplified from 25 ng genomic DNA using the primer sequences shown below:-

TG strand 5' ACA TAT GTA GTC TTC TGC AGG 3'

AC strand 5' ACA AGA GTG AAT CTC TGA CAG 3'

1 μCi [α-^{32}P]-dCTP (~6000 Ci/mmol; Amersham) was added to each PCR reaction, in a total volume of 25 μl. PCR was carried out for 29 cycles with 1 min at 94°C, 1 min at 55°C, and 1 min at 72°C, followed by 1 cycle with 1 min at 94°C, 1 min at 55°C, and 10 min at 72°C.

4 μl PCR product was mixed with 4 μl stop dye (95% formamide, 5% TBE, 0.1% xylene cyanol, 0.1% bromophenol blue), and 6 μl of this was then used to load the gel. Samples were run on 6% polyacrylamide (20:1; acrylamide:bisacrylamide) / 46% urea denaturing sequencing gels in a vertical electrophoresis unit (BRL) in 1 x TBE running buffer (90 mM Tris-borate, 2mM EDTA). Prior to loading, the samples were denatured at 95°C for 10 minutes and cooled on ice. Gels were run at a constant power of 70 W until the bromophenol blue dye had just run off the end of the gel (approximately 2 hours). Gels were then dried on a horizontal slab gel drier (BioRad) and exposed to Kodak XAR-5 X-Omat X-ray film in a cassette with double intensifier screens for 48 hours prior to automated developing. Individuals from family PK 2 were typed by Dr. P. Harris (Oxford), and some individuals from this family were loaded onto each gel to act as size markers.

2.2.4 PCR primers and conditions for YAC library screening

Each pair of primers was tested against total human genomic DNA, ThyB1-33-12/3 (ThyB1) mouse DNA and yeast DNA (AB1380) prior to screening of the libraries. None of the primer pairs gave a product when used on AB1380 alone.

2.2.4.1 PCR amplification of GGG1 (D16S259)

Primer sequences and starting conditions were supplied by Dr. M. H. Breuning (Leiden). The forward (M3) and reverse (R2) primer sequences are as follows:-

R2: 5' TCT GGG AAA GGA AGG GGT GGG CTC 3'

M3: 5' GTT GGG CAG GGC AGG GCG TGG TAG 3'

PCR was carried out for 1 cycle with 2 min 30 sec at 94°C, 1 min at 68°C and 3 min at 72 °C, followed by 27 cycles with 1 min at 94°C, 1 min at 68°C and 3 min at 72°C. The PCR product is approximately 1 kb in size, and samples were run on a 1% agarose gel in 0.5 x TBE using a horizontal electrophoresis tank. No product could be obtained using ThyB1 as a template.

2.2.4.2 PCR amplification of 26-6 (D16S125)

PCR was carried out as in section 2.2.2. The annealing temperature was reduced to 60°C for screening the mouse YAC library. The PCR product is 430 base pairs in size, and samples were identified by electrophoresis on a 3% Nusieve:1% LMP agarose gel.

2.2.4.3 PCR amplification of SM33p22

Primer sequences and starting conditions were supplied by Dr. M. H. Breuning (Leiden). The forward (148L) and reverse (147L) primer sequences are as follows:-

148L 5' GCC CCC GGC CCC GCG CGC CCC GCC CTG AGG AGC AGG GAG CAG AGG AAA C 3'

147L 5' ATG CGT GGC GTA TTC CTC TGG G 3'

PCR was carried out for 2 min at 94°C, followed by 28 cycles with 1 min at 94°C and 2 min at 66°C, finishing with 5 min at 72°C. The annealing temperature was reduced to 60°C for screening the mouse YAC library. The PCR product is 360 bp in length and was identified by electrophoresis on a 3% Nusieve:1% LMP agarose gel.

2.2.4.4 PCR amplification of ATPL

As described in Chapter 5, forward (C18) and reverse (C19) primers were designed to the 3' untranslated region of the sequence for AJ1 cDNA, an ATPase-like vacuolar proton channel gene which is located approximately 50 kb distal to 26-6DIS (Gillespie *et al.*, 1991). The primer sequences are as follows:-

C18 5' CCC GCC CCC AGT AGT TGG TC 3' (Tm = 68.1°C)

C19 5' AAC AGG AAC CCA CAA GGC GAC A 3' (Tm = 68.2°C)

PCR was carried out for 1 cycle with 2 min 30 sec at 94°C, 30 sec at 63°C and 1 min 30 sec at 72°C; followed by 25 cycles with 1 min at 94°C, 30 sec at 63°C and 1 min 30 sec at 72°C plus 3 sec extension per cycle; finishing with 1 cycle of 1 min at 94°C, 30 sec at 63°C and 5 min at 72°C. The annealing temperature was reduced to 57°C for screening the mouse YAC library. The PCR product is 295 bp in size, and was identified by electrophoresis on a 3% Nusieve:1% LMP agarose gel.

2.2.4.5 PCR amplification of VK5 (D16S94)

VK5 (D16S94) is a 7 kb *Eco*RI-*Bam*HI fragment in pBR322 (Breuning *et al.*, 1990a; Hyland *et al.*, 1990). As described in chapter 5, forward (VF1) and reverse (VR1) PCR primers were designed, thereby allowing PCR amplification of a 1.3 kb segment of VK5 (D16S94). The primer sequences are as follows:-

VF1 5' GGC CGC GTG TTT TGC TTT TG 3'

VR1 5' AGA AGG CAA CGA GGG CTG GC 3'

The $MgCl_2$ concentration in each reaction was 1.0 mM. PCR was carried out for 1 cycle with 2 min 30 sec at 94°C, 30 sec at 63°C and 1 min 30 sec at 72°C; followed by 25 cycles with 1 min at 94°C, 30 sec at 63°C and 1 min 30 sec at 72°C plus 3 sec extension per cycle; finishing with 1 cycle of 1 min at 94°C, 30 sec at 63°C and 5 min at 72°C. The annealing temperature was reduced to 60°C, and the $MgCl_2$ concentration increased to 1.5 mM for screening the mouse YAC library. The PCR product is 1.3 kb in length, and was separated on a 1% agarose gel.

2.2.5 *Alu* PCR

Primers used for *Alu* PCR were:-

TC-65 5' AAG TCG CGG CCG CTT GCA GTG AGC CGA GAT 3'

517 5' CGA CCT CGA GAT CTY RGC TCA CTG CAA 3' (where Y is a pyrimidine and R is a purine).

153 5' GGG ATT ACA GGC GTG AGC CAC 3'

154 5' TGC ACT CCA GCC TGG GCA 3'

450 5' AAA GTG CTG GGA TTA CAG G 3'

451 5' GTG AGC CGA GAT CGC GCC ACT GCA CT 3'

TC-65 and 517 are from Nelson *et al.* (1989); 153, 154, 450 and 451 from Breen *et al.* (1992). PCR was carried out as follows: 1 cycle with 3 min at 94°C, 1 min at 60°C and 1 min at 72°C, followed by 35 cycles with 45 sec at 92°C, 1 min at 60°C and 1 min at 72°C plus 6 sec extension per cycle, and finishing with 1 cycle of 45 sec at 92°C, 1 min at 60°C and 10 min at 72°C.

2.2.6 PCR amplification of catch-linkered YAC material

962L 5' GTC AAG AAT TCG GTA CCG TCG AC 3'

Catch-linkered material was amplified using the 962L primer under the usual reaction conditions, except that the primer concentration was 0.8 µM. PCR was carried out for 1 cycle with 2 min 30 sec at 94°C, 30 sec at 64°C and 2 min at 72°C, followed by 35 cycles with 30 sec at 94°C, 30 sec at 64°C and 2 min at 72°C, and finishing with 1 cycle of 30 sec at 94°C, 30 sec at 64°C and 5 min at 72°C.

2.2.7 PCR of YAC plugs

Following treatment with PMSF, 400 µl TE8 was added to each plug, and the plug melted at 65°C for 15 minutes. 1-5 µl of this was used in each PCR reaction.

2.2.8 PCR of YAC colonies

After Ross *et al.* (1992). A small amount of colony was picked into 20 µl lysis buffer (1% Triton X-100, 20 mM Tris HCl pH 8.3, 2 mM EDTA) and incubated at 100°C for 10 minutes. Cell debris was pelleted at 12,000 rpm in a microfuge for 5 minutes, and 5 µl of the supernatant used per PCR reaction.

2.2.9 PCR of recombinant DH5α colonies

Using a toothpick, a small amount of colony was placed in 20 µl sterile deionised water, vortexed, 20 µl phenol:chloroform:isoamyl alcohol (25:24:1 v/v) added,

vortexed again, and spun briefly in an Eppendorf centrifuge. 100 µl Phase Lock Gel™ (5Prime→3Prime Inc.) was added, the tube spun for 5 minutes and the supernatant removed. 1 µl of the supernatant was used in a PCR reaction with M13 universal and reverse primers.

2.3 PREPARATION OF DNA

2.3.1 Small scale preparation of plasmid and cosmid DNA (minipreps)

Rapid boiling method, after Holmes and Quigley (1981).

Cells from a 10 ml overnight L-broth cultures were harvested by centrifugation at 3,000 rpm for 10 minutes in a Sorvall RT6000B centrifuge, resuspended in 500 µl STET (8% w/v sucrose, 10 mM Tris-HCl, 50 mM Na_2EDTA, 0.5% v/v Triton X-100, pH 8.0) and transferred to a 1.5 ml microfuge tube. After addition of 15 µl lysozyme solution (10mg/ml lysozyme in STET), the tube was placed in a boiling water bath for 40 seconds. The lysate was then centrifuged for 10 minutes at 12,000 rpm, and the supernatant decanted into a clean tube. The volume of the supernatant was made up to 0.5 ml with sterile deionised water and extracted with an equal volume of phenol chloroform: isoamyl alcohol (25:24:1 v/v). The aqueous layer was removed and extracted with an equal volume of chloroform:isoamyl alcohol (24:1 v/v). The aqueous layer was removed, and the nucleic acids were precipitated with 0.1 volumes 3M sodium acetate (pH 6.0) and 0.6 volumes isopropanol, and collected by centrifugation at 12,000 rpm for 20 minutes at 4°C. After a 70% ethanol wash, the resulting pellet was dried and resuspended in 30 µl TE8 + RNAse A (10 µg/ml).
1-5 µl of this miniprep solution was usually sufficient for restriction enzyme analysis.

2.3.2 Preparation of DNA for sequencing

Cells from a 3 ml overnight Terrific broth (TB) culture were harvested by centrifugation at 12,000 rpm in an Eppendorf centrifuge for 5 minutes, and resuspended in 200 µl TE/0.1M NaCl. An equal volume of phenol:chloroform: isoamyl alcohol (25:24:1 v/v) was added, the microfuge tube vortexed hard, and

then centrifuged at 12,000 rpm for 7 minutes. The aqueous phase was removed, and 5 M ammonium acetate added to a final concentration of 2 M. DNA was collected by centrifugation at 12,000 rpm for 30 minutes, after precipitation with 1 ml ice-cold ethanol. The resulting pellet was washed with 70% ethanol, dried, and resuspended in 15 µl TE8 + RNAse A (10 µg/ml). 7-10 µl of this solution was used per sequencing reaction.

2.3.3 Preparation of yeast DNA plugs

10 ml of Casamino broth (-URA, -TRP) was inoculated with a loopful of yeast which had previously been streaked out on to a Casamino agar (-URA, -TRP) plate. The culture was incubated at 30°C with shaking at 200 rpm overnight, added to 90 ml YPD broth and incubated under the same conditions for a further 24 hours. The culture was decanted into two 50 ml Falcon tubes and the cells harvested by centrifugation in a Sorvall RT6000B at 3000 rpm (take up to speed and then stop). The cells were washed once with 50 mM EDTA pH 7.5 under the same centrifuge conditions and resuspended in 6 ml CPES (40 mM Citric acid, 120 mM Na_2HPO_4 pH 6.0, 1.2 M sorbitol dissolved together, pH adjusted to 6.0, made up to volume and autoclaved. 20 mM EDTA pH 7.5 added from autoclaved stock, and 5mM dithiothreitol (DTT) added from frozen 1 M stock just before use).

40 mg of Novo Zym 234 (Novo Industries AS, Denmark) was added to 10 ml 1% low melting-point (LMP) agarose in CPE (as CPES but without sorbitol or DTT) held at 38-42°C, and mixed well. This was added to the cells/CPES, mixed by inversion and placed in a water bath at 42°C to prevent premature solidification. 100 µl aliquots were pipetted into block formers (stored in 0.1 M HCl and washed with copious amounts of water prior to use) which had been sealed with tape and placed on ice. After setting, the plugs were removed and incubated at 30°C for 1 hour in CPE (about 1 ml CPE per block without shaking). The CPE was replaced with 1 ml/plug of Solution 3 (0.45 M EDTA pH 9.0, 10 mM Tris HCl pH 8.0, 1% sodium lauryl sarcosyl, 1 mg/ml proteinase K; made up from stocks just before use)

and the plugs incubated overnight at 50°C with gentle shaking. The CPE was then removed and the blocks stored at 4°C in 0.5 M EDTA pH 9.0.

Prior to use for PCR or restriction enzyme digestion, plugs were washed x4 in TE buffer pH 8.0 and then treated with PMSF. A 1000 x stock was made up by dissolving 40 mg PMSF in 1 ml 98% ethanol, and then a working solution was made of 0.04 mg/ml PMSF in TE8. Plugs were incubated in this solution at 50°C for 30 minutes (1 plug/ml) with gentle shaking, washed four times with TE8 and stored in TE8 until required for use (short-term storage only).

2.4 DNA MANIPULATIONS

2.4.1 Digestion of DNA

All enzymes used were obtained from Boehringer Mannheim unless otherwise stated. Digests were performed in a total volume of 20 µl, with 2 µl appropriate 10 x restriction enzyme buffer and an appropriate amount of enzyme. YAC plugs were treated with PMSF, washed in TE8 and equilibrated with 1 x buffer prior to the addition of enzyme.

2.4.2 Ligation reactions

Unless otherwise stated, ligations were performed in 10 µl total volume using 1 unit T_4 DNA ligase (Boehringer Mannheim), 100 ng digested and phosphatased pBluescript and an excess of insert. Ligation was carried out overnight, at 12°C for fragments with cohesive ends and at room temperature for fragments with blunt ends, using the manufacturers buffer.

2.4.3 Preparation of PCR fragments for subcloning

50 µl PCR product was removed and *Taq* polymerase inactivated by heating to 95°C for 15 minutes, and then cooling to room temperature to renature the PCR products. In order to create "flush ends", 1 µl Klenow enzyme (Boehringer Mannheim) and 5 µl 2 mM dNTPs were added and the mixture incubated at 37°C for 15 minutes. The PCR product was purified using Geneclean™ (Bio 101 Inc.),

according to the manufacturer's instructions, and resuspended in 10 µl TE8. 2 µl of product was electrophoresed in an agarose gel to check the concentration. 1 µl Polynucleotide kinase (10 units/µl; Boehringer Mannheim) was added to 0.5 µg product, and 10 x ligase buffer to a final concentration of 1 x. Samples were incubated at 37°C for 30 minutes and then 65°C for 15 minutes to inactivate the kinase.

2.4.4 Phosphatasing of pBluescript

pBluescript was digested with the appropriate enzyme to produce either compatible cohesive ends or blunt ends, and the number of picomole (pmol) ends calculated. 1 unit of bacterial alkaline phosphatase (BAP) was used per 2 pmol of ends. Tris HCl pH 8.0 was added to a final concentration of 50 mM, and the solution incubated at 37°C for 15 minutes. A further aliquot of BAP was added and incubated at 55°C for 45 minutes. SDS was added to a final concentration of 0.5%, EDTA to 5 mM and proteinase K to 100 µg/ml, and the tube incubated at 50°C for 30 minutes. DNA was extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 v/v) then with an equal volume of chloroform:isoamyl alcohol (24:1 v/v), and precipitated with 0.1 volumes 3 M sodium acetate pH 7.0 and 2 volumes ethanol. DNA was collected and washed as for miniprep DNA, and resuspended in TE pH 7.5 at a final concentration of 100 µg/ml. The vector was stored in aliquots at -20°C.

2.4.5 DNA sequencing

Sequencing reactions were performed using the Sequenase 2.0 DNA sequencing kit (USB). Plasmid DNA was prepared according to the protocol in section 2.3.2. For each plasmid, 10 µl DNA was added to 10 µl distilled water and 20 µl 2 x denaturing buffer (0.4 M NaOH, 0.4 M EDTA), and incubated at 37°C for 30 minutes. To this was added 4 µl 3 M sodium acetate and 88 µl absolute ethanol, and the denatured DNA precipitated at -70°C for 15 minutes. The DNA was pelleted by centrifugation for 20 minutes, washed with 70% ethanol and dried in a vacuum dessicator. The pellet was resuspended in 7 µl distilled water, 2 µl reaction buffer and 1 µl primer.

The primer was annealed at 37°C for 20 minutes, and then the Sequenase protocol was followed exactly for the labelling reactions.

Samples were run on 6% polyacrylamide / 46% urea denaturing sequencing gels in a vertical electrophoresis unit (BRL) in 1 x TBE running buffer (90 mM Tris-borate, 2 mM EDTA). Prior to loading, the samples were denatured at 95°C for 3 minutes and cooled on ice. Gels were run at a constant power of 70 W for 1½ hours for the short run and 4 hours for the long run. Gels were then dried on a horizontal slab gel drier (BioRad) and exposed to Kodak XAR-5 X-Omat X-ray film in a cassette with double intensifier screens for 36 hours.

2.5 TRANSFORMATION

2.5.1 Preparation of competent cells

A 10 ml culture of DH5α in L-broth was grown overnight, 0.5 ml of this was added to a fresh 10 ml of L-broth and grown for 1½ hours. The cells were spun down, the supernatant removed and the cells resuspended in 1 ml ice-cold 100 mM $CaCl_2$ (filter-sterilised) and placed on ice for 10 minutes. The above steps were repeated, the cells spun a third time and resuspended in 200 µl 100 mM $CaCl_2$ ready for transformation.

2.5.2 Transformation of DH5α

200 µl competent DH5α cells were made fresh (see 2.6.1) or thawed (Gibco BRL), and left on ice for 10 minutes. 3-5 µl of ligation mix was added and the tube left on ice a further 30 minutes. After heat shock at 42°C for 90 sec, the tube was placed on ice for 2 minutes. Three volumes of SOC were added and the cells incubated with shaking at 37°C for 1 hour, spun down for 30 sec and resuspended in 200 µl SOC. 100 µl of cells were plated on L-agar plates spread with 80 µl ampicillin (25 mg/ml), 40 µl 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal, 20 mg/ml in dimethylformamide) and 4 µl isopropylthio-β-D-galactoside (IPTG, 200 mg/ml in H_2O, filter-sterilised) and incubated at 37°C overnight.

2.6 SOUTHERN TRANSFER

After electrophoresis in agarose gels, DNA was transferred to Hybond-N^+ membranes (Amersham) by the alkali blotting method as described in the Hybond-N^+ protocol booklet. A capillary blot was set up using 0.4 M NaOH as the transfer buffer, and transfer allowed to proceed overnight. The filter was rinsed in 2 x SSC, and was then ready for hybridisation.

2.7 HYBRIDISATION PROTOCOL

All hybridisations were based on the following protocol, and any changes are noted in the relevant sections.

Filters were prehybridised at 65°C for 3 hours in double sealed plastic bags with 10-20 ml hybridisation fluid [7% (w/v) sodium dodecylsulphate (SDS), 0.5 M Na_2PO_4 pH 7.4, 1% (w/v) bovine serum albumin (BSA)] and 0.1 mg/ml denatured salmon sperm DNA. Probes were labelled with [α-^{32}P]-dCTP (~6000 Ci/mmol) using a random priming kit (Boehringer Mannheim), and unincorporated counts were separated by gel filtration. Filters were then hybridised in the same mixture containing 1 x 10^6 cpm/ml (1-3 ng/ml) of labelled probe. After overnight hybridisation at 65°C, filters were washed in 2 x SSC/0.1% SDS down to a final stringency of 0.1 x SSC/0.1% SDS at 65°C. Filters were used to expose Kodak XAR-5 films in cassettes with double intensifier screens for 1-14 days at -70°C.

Filters were stripped of probe by boiling for 10 minutes in 0.5% (w/v) SDS and allowing to cool to room temperature.

2.8 SCREENING OF YAC LIBRARIES

The YAC libraries were primarily screened using the polymerase chain reaction (PCR), based on Green and Olson (1990), with the sensitivity increased by Southern blotting the gels on which the PCR products were separated, and probing the filters. Some of the CEPH library secondary screening was performed by colony hybridisation. PCR primers and conditions are described in section 2.2.4.

2.8.1 CEPH YAC library (Albertsen *et al.*, 1990)

This library was obtained from CEPH with the kind permission of Dr. D. Le Paslier. The 113 master pools were sent as 500 ng lyophilised DNA which was resuspended in 50 µl ultra-pure sterile distilled water. 10 ng (1 µl) was used for each PCR reaction: all reactions were carried out in a total volume of 10 µl using the primers and conditions detailed below. Each master pool contained 384 clones (4 x 96-well plates), and secondary screening was performed by colony hybridisation. Later in the project, subpools consisting of rows x columns became available for secondary screening and 5 µl of this DNA was used in each 25 µl PCR reaction. The CEPH YAC library was screened using primers to the GGG1 (D16S259), ATPL and VK5 (D16S94) loci.

2.8.2 ICI YAC library (Anand *et al.*, 1990)

This library was initially obtained from Dr. R. Anand (ICI Diagnostics), and the latter stages of screening were performed with the help of Dr. R. Elaswarapu (HGMP Resource Centre). The library consists of 40 master pools each representing 9 subpools, and each subpool represents a 96-well microtitre plate. The master pools and subpools were obtained as DNA plugs, and these were washed in TE buffer under sterile conditions. 500 µl TE was added to each plug which was then melted at 65°C. 2 µl of plug DNA was used in each PCR reaction. Following the identification of positive subpools, rows x columns for the relevant 96-well plate were sent as a cell suspension from the HGMP Resource Centre. 3 µl of the cell suspension was used in each PCR reaction. All reactions were carried out in a total volume of 50 µl. The ICI YAC library was screened using primers to the GGG1 (D16S259), 26-6 (D16S125), SM33p22, ATPL and VK5 (D16S94) loci.

2.8.3 St. Mary's Hospital Medical School mouse YAC library (Chartier *et al.*, 1992)

This library was obtained by kind permission of Dr. S. D. M. Brown, and distributed by Dr. F. L. Chartier. The library consists of 16 complex pools, each corresponding

to 1920 different clones made up from 10 simple pools. The simple pools correspond to 2 microtitre plates, and the final set of screening is done by preparing rows x columns of pools in individual microtitre plates. 5 µl of DNA per 50 µl reaction was used in screening of the complex pools, and 2 µl of DNA per 25 µl reaction for the simple pools. Prior to the preparation of rows x columns PCR pools, the microtitre plates were duplicated using a sterile 96-pronged replicator. The YAC colonies were grown up for 24-48 hours. Twenty-eight Eppendorf tubes were labelled, 1-12 for columns and A-P for rows. Under sterile conditions, 200 µl YPD broth was added to each well and the yeast resuspended. 80 µl was then distributed into the row tube, and 80 µl into the column tube. The tubes were spun for 5 minutes in an Eppendorf centrifuge, the supernatant removed, the cells resuspended in PCR quality water, and the tubes spun again. Finally the cells were resuspended in 200 µl lysis buffer (50 mM KCl, 10 mM Tris HCl pH 8.0, 1.4 mM $MgCl_2$, 0.01% gelatin, 0.45% Tween-20, 0.45% NP-40, 1 U/µl yeast lytic enzyme. The Tween-20 and NP-40 are added after autoclaving, and the yeast lytic enzyme just prior to use), transferred to PCR tubes and incubated at 37°C for at least 1 hour, followed by 100°C for 10 minutes. 2 µl of each pool was used per 25 µl PCR reaction. This mouse YAC library was screened with primers to the 26-6 (D16S125), SM33p22, ATPL and VK5 (D16S94) loci. The primers were tested on ThyB1 (mouse) DNA, and annealing temperatures adjusted as required, prior to screening of the library.

2.8.4 Probing of PCR products

The PCR products were electrophoresed in agarose gels along with a control sample of genomic DNA that had been PCR amplified at the same time. DNA was transferred to Hybond-N^+, and filters prehybridised. For each locus, genomic DNA was PCR amplified, the product purified using Geneclean™ (Bio 101 Inc.), and diluted to 25 ng/µl for use as a probe. Filters were hybridised as per the protocol and used to expose Kodak XAR-5 films in cassettes for 1-7 days at -70°C.

2.8.5 Secondary screening by colony hybridisation

14 cm diameter petri dishes were poured using Casamino agar (-URA, -TRP) with added ampicillin (100 µg/ml). Each 96-well plate was inoculated onto one petri dish using a sterile toothpick and grid. Colonies were grown to 1 mm in diameter by incubation at 30°C (approximately 36 hours). Nytran filters (Schleicher and Schuell) were labelled after sterilisation by autoclaving, placed over the colonies and orientated. Colonies were transferred by pressure using a sterile glass spreader, and the filter was then placed colony side up on another agar plate. Both master and replica plates were incubated at 30°C until the colonies were of the appropriate size. Once the master plate had recovered, a second replica lift was performed.

The filters were then prepared for hybridisation as per the following protocol. Filters were placed in large square petri dishes, colony side up on 3MM Whatmann paper soaked in SCE (1 M sorbitol, 600 mM EDTA, 0.1 M sodium citrate) with 4 mg/ml Novo Zym and 10 mM DTT. The petri dishes were sealed in polythene to avoid evaporation, and incubated at 37°C overnight. Filters were then placed on 3MM Whatmann paper soaked in denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 20 minutes, removed and dried for 10 minutes, floated colony side up on 1 M Tris HCl pH 7.6/1.5 M NaCl for 5 minutes, then 0.1 M Tris HCl /0.15 M NaCl for 5 minutes. Filters were then submerged in 0.1 M Tris HCl /0.15 M NaCl with 250 µg/ml proteinase K (Boehringer Mannheim) and incubated at 37°C for 60 minutes. Any remaining debris was wiped off the filters with a soft tissue soaked in 0.1 M Tris HCl /0.15 M NaCl, prior to submerging in 50 mM $Na_2HPO_4.2H_2O$ pH 7.2, drying, and baking at 80°C for 30 minutes. Hybridisation was performed as in the usual protocol, and filters used to expose Kodak XAR-5 films in cassettes for 12 days at -70°C.

2.9 PULSED FIELD GEL ELECTROPHORESIS (PFGE)

All PFGE was done in a Pharmacia LKB 2015 CHEF electrophoresis unit using the LKB 2015 pulsaphor plus control unit and LKB 2301 macrodrive 1 power supply, cooled with a LKB Multitemp II thermostatic circulator.

2.9.1 Separation of yeast chromosomes

Whole YAC plugs were placed on the comb, and a 1% agarose gel poured around them so that they were fully submerged. AB1380 plugs were used as a yeast control. The chromosomes were separated using the following parameters:-

- 165v throughout
- N/S 70 sec, E/W 70 sec for 21 hours followed by N/S 120 sec, E/W 120 sec for 21 hours
- running temperature 8°C

The gel was visualised by ethidium bromide staining and treated with 0.25 M HCl for 15 minutes. The DNA was transferred to a Hybond-N^+ filter using the alkali blotting technique as above. Probes were used without purification of inserts [CMM65 (D16S84), EKMDA (D16S83) and 218EP6 (D16S246)] and labelled with [α-^{32}P]-dCTP by random priming. Hybridisation was carried out exactly as for probing of PCR products above. Filters were used to expose Kodak XAR-5 films in cassettes with double intensifier screens overnight at -70°C.

2.9.2 Isolation of YACs

A 1% agarose gel was poured, the centre cut out, and YAC plugs were stacked 3 high against the comb and a 1% LMP agarose gel poured around them. The YACs were separated from the larger yeast chromosomes and run into the LMP agarose using the following programme:-

- 165v throughout
- N/S 30 sec, E/W 30 sec for 24 hours
- running temperature 4°C

The gel was visualised by ethidium bromide staining, the YAC bands cut out using a sterile scalpel blade and stored in TE8.

2.9.3 Separation of YAC digests and preliminary mapping

PMSF treated plugs were taken and 1 plug per Eppendorf tube was equilibrated with 200 µl 1 x restriction enzyme buffer on ice for 30 minutes. The buffer was

removed and replaced with 100 µl fresh buffer plus 3 µl restriction enzyme. Plugs were digested overnight at the temperature appropriate to each enzyme. After cooling on ice, plugs were cut in half and loaded on a 0.9% agarose gel. The fragments were separated by PFGE using the following program:-

- 165v throughout
- N/S 2 sec, E/W 2 sec for 12 hours followed by N/S 5 sec, E/W 5 sec for 12 hours
- running temperature 8°C

The gel was visualised by ethidium bromide staining and treated with 0.25 M HCl for 15 minutes. The DNA was transferred to a Hybond-N^+ filter using the alkali blotting technique, as above. The filter was probed with pBluescript as the YAC left arm probe, stripped and re-probed with a 774 bp fragment of pBR322 as the YAC right arm probe. The right arm probe was made by digesting pBR322 with *Ava*I followed by *Sal*I, electrophoresing the products into a 1.5% agarose gel, excising the 774 bp fragment, treating with agarase, precipitating and resuspending the DNA. Probes were labelled by random priming and hybridisation was carried out as per the usual protocol. Filters were used to expose Kodak XAR-5 films in cassettes with double intensifier screens for 3 days at -70°C.

2.10 ATTACHMENT OF CATCH-LINKERS

Blunt-end and cohesive-end catch-linkers, along with the complementary primer to allow PCR amplification, were designed by Dr. J. Brown (MRC, HGU).

961L 5' GTC GAC GGT ACC GAA TTC 3'

963L 5' GAT CGT CGA CGG TAC CGA ATT CT 3'

962L 5' GTC AAG AAT TCG GTA CCG TCG AC 3'

961L is the blunt-end complement of master primer 962L, and 963L is the *Sau*3A compatible complement of master primer 962L. The linkers are made by making a solution of either 25 µM 961L/25 µM 962L (MF blunt-end linker) or 25 µM 963L/ 25 µM 962L (963L *Sau*3A compatible linker), boiling for 10 minutes and annealing by cooling slowly to room temperature. The stock solution is kept at -20°C.

The YAC band was excised from LMP agarose after PFGE as described above (approx. 200 µl of agarose) and equilibrated with 1 ml restriction enzyme buffer for 1 hour at room temperature. The buffer was then removed and replaced with 250 µl buffer plus 2 µl restriction enzyme (8-10 units/µl), either *Bgl*II or *Rsa*I. *Bgl*II gives *Sau*3A-compatible cohesive ends, and *Rsa*I blunt ends. The plug was incubated at 37°C for 4 hours and then the enzyme inactivated and the plug melted by incubation at 65°C for 10 minutes. 10 x Agarase buffer was added to a final concentration of 1 x, along with 4 µl β-agarase (1 unit/µl; New England Biolabs) and the tube incubated at 40°C for 1 hour. The DNA was extracted with an equal volume of phenol: chloroform:isoamyl alcohol (25:24:1 v/v), the aqueous layer removed and extracted with an equal volume of chloroform:isoamyl alcohol (24:1 v/v). The aqueous layer was removed, 0.1 volumes 3 M sodium acetate (pH 6.0) was added, the tube left on ice for 15 minutes and then spun in an Eppendorf centrifuge at 12,000 rpm for 15 minutes at 4°C. The supernatant was removed, 2 volumes of ethanol and 1 µl glycogen (20 mg/ml; Boehringer Mannheim) added and the DNA precipitated overnight at -70°C. The DNA was collected by centrifugation for 30 minutes, the resulting pellet washed with 70% ethanol, air-dried and resuspended in 10 µl TE8.

Ligation reactions were set up using 5 µl restricted YAC DNA, 3 µl linkers (either 25 µM or 2.5 µM), 1 µl T_4 DNA ligase (Boehringer Mannheim) and 1 µl 10 x ligation buffer. DNA restricted with *Bgl*II was ligated to the cohesive-end linker 962L/963L, and DNA restricted with *Rsa*I was ligated to the blunt-end linker MF. Cohesive-end ligations were allowed to proceed overnight at 12°C and blunt-end ligations at room temperature. 90 µl of sterile distilled water was added to each tube, and the reactions stopped by incubating at 65°C for 15 minutes.

2.11 FLUORESCENCE IN SITU HYBRIDISATION

This was carried out by Dr. M. Breen (YACs 71B2 and 71B4), Ms. S. Boyle (YAC 32A-A4) and Mrs. M. Harris(YAC C4-I12). Approximately 1 µg of amplified catch-linkered YAC material was used for *in situ* hybridisation. Metaphase spreads were

prepared from phytohaemagglutinin stimulated peripheral blood lymphocyte cultures or lymphoblastoid cell cultures (BB33) by routine methods of colcemid, potassium chloride hypotonic and methanol/acetic acid fixation treatments. Slide preparations were air dried and stored under vacuum for up to two weeks. The PCR amplified YAC fragments were labelled with biotin-16-dUTP (Boehringer Mannheim) by a nick translation reaction (Rigby *et al.*, 1977) in which the concentration of DNAse I was adjusted to give fragment sizes averaging 200-500 bp. Labelled probe was separated from the reaction by passing through a Nick™ column (Pharmacia) equilibrated in 0.01 M Tris HCl pH 8.0 / 0.2 M NaCl / 0.01 M EDTA. The level of biotin incorporation was determined using an avidin alkaline phosphatase-NBT/BCIP substrate assay (Vector laboratories).

For each slide, 50 ng of biotinylated amplimered YAC, 5 µg human *Cot1* DNA (BRL) and 5 µg sonicated salmon sperm DNA were mixed and precipitated with absolute ethanol. The DNA was pelleted by microcentrifugation and resuspended in 10 µl of a hybridisation mixture containing 50% formamide, 10% dextran sulphate, 2 x SSC (0.3 M NaCl; 0.03 M Tri-sodium citrate), 0.1% Tween-20. The DNA was denatured at 75°C for ten minutes and then pre-annealed at 37°C for 30 minutes. Immediately before use, slide preparations were passed through an ethanol series (70%-90%-100%). Chromosomal DNA was denatured at 70°C in 70% v/v formamide/2 x SSC for two minutes and then passed through a cold ethanol series (70%-90%-100%). Hybridisation was performed under a sealed 22 x 22 mm coverslip at 37°C for 16 hours. Post-hybridisation washing was carried out according to Pinkel *et al.* (1986) except that cytochemical detection was as described by Gosden *et al.* (1991). Slides were mounted in antifade (1:1 glycerol: Citifluor AF3) containing 3.75 µg/ml DAPI and 0.75 µg/ml propidium iodide. This mixture of fluorochromes produces an R-banded pattern on the majority of chromosome preparations treated as described above. Analysis of results was carried out using a Leitz Ortholux II microscope fitted with a Ploemopak filter system. Suitable chromosome preparations were scanned with a BioRad Lasersharp

MRC 600 confocal scanner attached to the same microscope. Map positions were determined by observation of the R-banding pattern.

2.12 END-CLONING TECHNIQUES

End-cloning of YACs 32A-A4 and C4-I12 was attempted using various techniques detailed below.

2.12.1 End-cloning by asymmetric PCR

After Rosenthal *et al.* (1991). Vector primers 373 (left) and 556 (right) were taken from Arveiler and Porteous (1991) and biotinylated as 496N (left) and 497N (right).

496N 5' Biotin-GCC AAG TTG GTT TAA GGC GC 3'

497N 5' Biotin-GCC CGA TCT CAA GAT TAC G 3'

20 µl PCR amplified catch-linkered YAC DNA was purified by Geneclean™ and resuspended in 100 µl TE8. Asymmetric PCR was performed on 1 µl template with 90 sec at 95°C, 50 cycles of 30 sec at 95°C, 1 min at 55°C, 1 min at 72°C, finishing with 3 min at 72°C. The biotin-labelled PCR products were isolated with Dynabeads™ (Dynal) as in Rosenthal *et al.* (1991), and the anchored DNA template amplified with primers 962L and either 373 (left end) or 556 (right end), using the same program as above with 35 cycles. Any product was re-amplified with the same primers and subcloned into pBluescript.

2.12.2 End-cloning by *Alu*-vector PCR

1 µl YAC plug DNA was used as the template. Vector primer 373 or 556 was added first and 5 cycles of asymmetric PCR performed before adding one of the *Alu* primers. The PCR program was as for *Alu*-PCR. Any product was re-amplified with the same primers and subcloned into pBluescript.

2.13 cDNA LIBRARIES

2.13.1 Adult kidney cDNA library

An adult kidney cDNA library was purchased from Clontech Laboratories. Poly A^+ RNA had been isolated from a young male whose kidney had been perfused for 24 hours. The library was made by oligo (dT) priming and cloned into the lambda gt10 (λgt10) vector at the *Eco*RI cloning site. The average insert size is 0.9 kb (range 0.5-3.2 kb). Titration was performed as recommended in the library handbook, and the titre found to be 2.35×10^{10} pfu/ml.

2.13.2 Foetal kidney cDNA library

This library was obtained from Dr. J. K. Cowell, ICRF. The library has an average insert size of 1.0 kb and was made by oligo (dT) priming and cloned into the lambda gt10 (λgt10) vector at the *Eco*RI cloning site. Titration was performed as for the adult library and the titre found to be 1.25×10^{11} pfu/ml.

2.14 YAC-cDNA HYBRIDISATION

2 µl of the adult kidney cDNA library was diluted 1:10 (2×10^6 pfu/µl), and 10 µl used per PCR reaction to achieve adequate representation. 2 µl of the foetal library was also diluted 1:10 (1×10^7 pfu/µl), and 10 µl used per PCR reaction. The diluted cDNA libraries were subjected to two freeze/thaw cycles prior to use.

2.14.1 Primary amplification of cDNA

The cDNA libraries were amplified using the λgt10 forward and reverse primers at a concentration of 0.8 µM each. Ten PCR reactions were set up for each of the libraries using 10 µl 1:10 diluted cDNA as template in a total volume of 50 µl. PCR was carried out for 1 cycle with 3 min at 94°C, 1 min at 62°C and 3 min at 72°C; 20 cycles with 1 min at 93°C, 1 min at 62°C and 3 min at 72°C and 1 cycle with 1 min at 93°C, 1 min at 62°C and 10 min at 72°C. The 10 reactions were then pooled, the mineral oil removed by chloroform extraction and the DNA purified using Geneclean™.

2.14.2 Primary amplification of YAC DNA.

Ten PCR reactions were set up using 1 µl of *Bgl*II-digested, catch-linkered YAC DNA. The PCR program used was as before. As with the cDNA, the 10 reactions were pooled, the mineral oil removed by chloroform extraction and the DNA purified using Geneclean™.

2.14.3 Secondary amplification

1 µl of DNA from the primary amplification step was used as a template for further PCR. Ten reactions were set up for both the cDNA and YAC DNA. The cDNA was amplified as above, but the extension times were increased to 7 minutes per cycle. The YAC DNA was amplified as above, but biotinylated 962L (C232) primers were used. Amplification products were again pooled and purified.

2.14.4 Hybridisation and selection of enriched cDNAs

Biotinylated YAC DNA was pre-competed with human *Cot*I DNA (BRL) and an excess of 962L primers (to prevent self-annealing of the C232 primers and consequent hairpin formation). At the same time, amplified cDNA was pre-competed with human *Cot*I DNA.

4 µl (600 ng) biotinylated YAC DNA, 5 µl 10 µM 962L and 20 µl 1 mg/ml human *Cot*I DNA (BRL) were mixed together and the DNA precipitated with 2.9 µl 3 M sodium acetate pH 6.0 and 90 µl absolute ethanol at -70°C for 30 minutes. The DNA was pelleted by centrifugation for 15 minutes at 4°C, washed with 70% ethanol, dried, resuspended in 50 µl hybridisation buffer (6 x SSC, 0.1% SDS, 1 mM EDTA, 10 mM Tris HCl pH 7.5) in a PCR tube and covered with 30 µl mineral oil. 40 µl (10 µg) amplified cDNA was mixed with 20 µl *Cot*I DNA, 6 µl sodium acetate and 180 µl ethanol were added and the DNA precipitated as above, resuspended in 50 µl hybridisation buffer and covered with 30 µl mineral oil. Both tubes were incubated at 100°C for 15 minutes, cooled to 68°C and incubated at 68°C for 2½ hours. The two tubes were mixed together giving a total volume of 50 µl and this was incubated at 68°C for 20 hours to allow hybridisation to occur.

50 µl Dynabeads™ (Dynal) were washed x 3 with binding/washing (B/W) buffer (10 mM Tris HCl pH 7.5, 1 mM EDTA, 2 M NaCl) then resuspended in 50 µl B/W buffer + 30 µg denatured salmon sperm DNA. The beads were incubated at room temperature for 15 minutes, washed with B/W buffer, resuspended in the 100 µl YAC/cDNA hybridisation mixture and incubated again at room temperature for 15 minutes. The beads were then washed x 5 with 0.1% SDS, 0.1 M NaCl, 1 mM EDTA at 68°C for 10 minutes per wash. After the final wash, the beads were resuspended in 50 µl 2 x PCR buffer (Promega) and transferred to a 0.5 ml Eppendorf tube for elution. The tube was heated to 75°C for 10 minutes, the supernatant removed and kept and replaced by a further 50 µl 2 x PCR buffer. The elution step was repeated at 85°C and 95°C, each time removing and keeping the supernatant. 50 µl distilled water was added to each supernatant to produce 1 x PCR buffer.

1 µl of this eluate was used as a template for PCR with the λgt10 forward and reverse primers as before. This amplified, enriched cDNA could then be taken through further rounds of hybridisation. Once discrete bands were seen in the enriched cDNA mixture, PCR of each band was performed by using a 23G needle to pick up a piece of the band from the gel and inoculate the PCR reaction mix. The bands were amplified with λgt10 primers for 20 cycles taking 5 µl of product into fresh PCR mix (to 50 µl final volume) and re-amplifying for a further 15 cycles. 15 µl of this product was again put into fresh PCR mix (to 50 µl final volume) and run through a "clean-up" program in order to reduce the concentration of partially extended and self-annealed products. This consisted of 1 cycle at 94°C for 4 min, 62°C for 2 min and 72°C for 10 min, followed by 2 cycles of 62°C for 2 min and 72°C for 10 min.

2.15 COSMID SCREENING

An aliquot of a mouse cosmid library (Poustka *et al.*, 1984) was kindly sent by Dr. J. H. McVey (Haemostasis Research Group, CRC, Harrow). The titre was confirmed to be 1 x 10^9/ml after dilution plating. The mouse genome is 3 x 10^9 base pairs,

and therefore approximately 150,000 cosmid colonies are required to provide two genome equivalents. 22 x 22 cm L-agar + 25 μg/ml kanamycin plates were prepared and Hybond-N^+ filters sterilised by autoclaving. 1 μl of the library (1 x 10^6 colonies) was diluted with 5 μl L-broth, and 1 μl of this dilution added to 20 ml L-broth (166,666 colonies). This was spread over the master filter using suction apparatus; the filter was then placed on an agar plate and incubated overnight at 37°C. Replica filters were made by removing the master filter onto sterile filter paper in a sterile environment, placing a second filter on top and using a rolling pin to transfer the colonies. The filters were orientated by a series of needle-stick holes, the filters were peeled apart, placed colony side up on agar plates and incubated until the colonies were 1 mm in size. A second replica filter was made in the same manner. The filters were prepared for hybridisation by placing on 3MM Whatmann paper soaked in the following solutions: 1.5 M NaCl/0.5 M NaOH for 7 minutes, 0.5 M Tris HCl pH 7.4/1.5 M NaCl for 5 minutes and fresh 0.5 M Tris HCl pH 7.4/1.5 M NaCl for 5 minutes. The filter was washed briefly in 2 x SSC and colony debris removed with a gloved hand. DNA was fixed to the membrane by placing on 3MM Whatmann paper soaked in 0.4 M NaOH for 20 minutes. Finally, the filter was rinsed briefly (maximum 1 minute) in 5 x SSC and dried prior to hybridisation.

Filters were prehybridised in 20 ml hybridisation fluid and hybridised using the above protocol. 100 ng amplified catch-linkered C4-I12 mouse YAC DNA was labelled by random priming and unincorporated counts removed by gel filtration. 700 μl probe (1 x 10^6 cpm/ml hybridisation fluid) was added to 300 μl 20 x SSC (5 x SSC final), 100 μl 1 μg/μl mouse *Cot*I DNA (BRL) and 100 μl TE8, boiled for 10 minutes, placed on ice for 1 minute and allowed to pre-compete at 65°C for 2 hours. 600 μl of this probe mix was added to each filter, and hybridisation continued overnight. Filters were used to expose Kodak XAR-5 films in cassettes with double intensifier screens for 10 days at -70°C. Replicate signals were identified and marked and the corresponding colonies on the master filter picked and resuspended in 100 μl L-broth + kanamycin. Serial dilutions were made, the 1:100 and 1:1000 dilutions plated onto Hybond-N^+ filters on L-agar + kanamycin plates,

and incubated at 37°C overnight. Replicas were made of the 1:1000 filters (approximately 100 colonies), and these were probed with C4-I12 as above.

2.16 CHEMICALS

Chemicals were purchased either from BDH Chemicals Ltd. or from Sigma Chemical Co. Ltd. unless otherwise stated. Media constituents were obtained from Difco Laboratories.

2.17 GROWTH MEDIA

L-broth/L-agar

bacto-tryptone	10 g
bacto-yeast extract	5 g
NaCl	10 g
(bacto-agar	20 g)

Add deionised water to 1 litre, and adjust pH to 7.0 with 5M NaOH. Sterilise by autoclaving for 20 minutes at 15 lb/sq. in. on liquid cycle.

Terrific broth (TB)

bacto-tryptone	12 g
bacto-yeast extract	24 g
glycerol	4 ml

Add deionised water to 900 ml, and sterilise by autoclaving for 20 minutes at 15 lb/sq. in. on liquid cycle.

Allow the solution to cool to 60°C or less, and then add 100 ml of a sterile solution of 0.17 M KH_2PO_4, 0.72 M K_2HPO_4. (This solution is made by dissolving 2.31 g of KH_2PO_4 and 12.54 g of K_2HPO_4 in 90 ml of deionised water, adjusting the volume to 100 ml with deionised water and sterilising by autoclaving for 20 minutes at 15 lb/sq. in. on liquid cycle.)

Casamino broth/agar (-URA, -TRP)

yeast nitrogen base without amino acids	1.7 g
casamino acids	5 g
adenine hemisulphate	0.008 g
(bacto-agar	20 g)

Add deionised water to 1 litre and pH to 5.8. Sterilise by autoclaving for 20 minutes at 15 lb/sq. in. on liquid cycle. After autoclaving, add 50 ml sterile 40% D-glucose.

YPD broth/agar

bacto-yeast extract	10 g
bacto-peptone	20 g
(bacto-agar	20 g)

Add deionised water to 1 litre, and pH to 5.8. Sterilise by autoclaving for 20 minutes at 15 lb/sq. in. on liquid cycle. After autoclaving, add 50 ml sterile 40% D-glucose.

SOC medium

bacto-tryptone	20 g
bacto-yeast extract	5 g
NaCl	0.5 g

Add the above to 950 ml deionised water and shake until dissolved. Add 10 ml 250 mM KCl, adjust the pH to 7.0 and the volume to 1 litre. Sterilise by autoclaving for 20 minutes at 15 lb/sq. in. on liquid cycle. After autoclaving, add 20 ml sterile 1 M sterile glucose. Just before use add 5 ml sterile 2 M $MgCl_2$.

2.18 BACTERIAL AND YEAST STRAINS

The genotype of the yeast host strain AB1380 is as follows:-

MATα ψ^+ *ura3 trp1 ade2-1 can1-100 lys2-1 his5*

The genotype of the host bacterial strain DH5α is as follows:-

supE44 hsdR17 recA1 endA1 gyrA96 thi-1 relA1

CHAPTER THREE

GENETIC MAPPING I

3.1 LINKAGE ANALYSIS AND GENETIC HETEROGENEITY

Following the initial demonstration of linkage between ADPKD and 3'HVR (D16S85) (Reeders *et al.*, 1985), many linkage studies have been performed on ADPKD kindreds. Several of these studies have shown no evidence of linkage heterogeneity (Reeders *et al.*, 1987; Lazarou *et al.*, 1987; Mandich *et al.*, 1990; Turco *et al.*, 1991) despite reports of individual large families with typical ADPKD which do not demonstrate linkage with chromosome 16p markers (Romeo *et al.*, 1988; Kimberling *et al.*, 1988; Brissenden *et al.*, 1989; Nørby and Schwartz, 1990). The above data would suggest that the number of PKD2 (unlinked) families is small in comparison with the number of PKD1 (linked) families. In the provision of ante-natal and pre-symptomatic diagnosis, and genetic counselling, it is important to establish the extent of genetic heterogeneity in each population. Ten Scottish ADPKD kindreds were included in the heterogeneity study by Reeders *et al.* (1987) which found no evidence for a second ADPKD locus in kindreds from four different countries (England, Scotland, Holland and Finland). The present study has extended the Scottish sample to a total of 35 families, in order to clarify the extent of genetic heterogeneity in this population, and to identify each family as either PKD1 or PKD2.

In the absence of cytogenetic abnormalities that mark the position of the disease gene, the correct identification of close flanking markers on either side of the disease mutation is critical, since this is the only means of defining the physical boundaries of the region containing the disease locus. Accurate localisation of the PKD1 gene requires a genetically homogeneous sample and a large number of kindreds. In order to facilitate accurate genetic mapping of PKD1, many markers have been isolated from the short arm of chromosome 16, and these have been ordered with respect to each other by a combination of linkage analysis, physical

mapping to human-rodent hybrids, and long-range restriction mapping (Reeders *et al.*, 1988; Callen *et al.*, 1989; Breuning *et al.*, 1990a; Germino *et al.*, 1990; Harris *et al.*, 1990). Several human-rodent hybrids are available, each containing an aberrant chromosome 16, deleted for various parts of 16p. They include CY13 (GM6227), 23HA (GM2324), N-OH1, and CY14 (Figure 3.1).

CY13 is a subclone of the 46,XX,t(1;16)(q44;p13.11) lymphoblastoid cell line 6227-1-C10. It contains the derivative chromosome 16qter-16p13.11, and is therefore deleted for the region 16p13.11-16pter (Callen *et al.*, 1986).

23HA contains the der16 chromosome from GM2324, which is an EBV transformed cell line from a reciprocal translocation carrier with the karyotype 46,XX,t(16;22)(p13.3;q12.2) (Germino *et al.*, 1990).

N-OH1 is a somatic cell hybrid containing the abnormal chromosome 16 from a proband with α-thalassaemia, mental retardation and a subtelomeric translocation. The proband has the karyotype 46,XY,-16,+der(16),t(1;16) (p36.3;p13.3), and is trisomic for the tip of chromosome 1 and monosomic for the tip of chromosome 16 (Lamb *et al.*, 1989).

CY14 is a mouse/human hybrid containing the aberrant chromosome 16 from a cell line with an unbalanced translocation 46,XY,-16,+der(16)t(4;16) (q31.1;p13.3) (Breuning *et al.*, 1987a; Callen *et al.*, 1989). Detailed maps of chromosome 16p have therefore been constructed, and a summary is shown in Figure 3.2.

3.2 ANALYSIS OF FAMILIES

Thirty-five families were ascertained through patients with a diagnosis of ADPKD attending the Medical Renal Unit, Royal Infirmary of Edinburgh, or the Renal Unit, Western Infirmary, Glasgow, as described in Materials and Methods. The families had previously been typed for 3'HVR (D16S85) and CMM65 (D16S84), and many had been typed for VK5 (D16S94). Most of this work was carried out by Ms. P.M. Pignatelli. Typing of the kindreds with VK5 (D16S94) was completed, and families were also typed with two new markers, 26-6 (D16S125) and SM7 (D16S283).

Figure 3.1 CHROMOSOME 16p13 HYBRID BREAKPOINTS

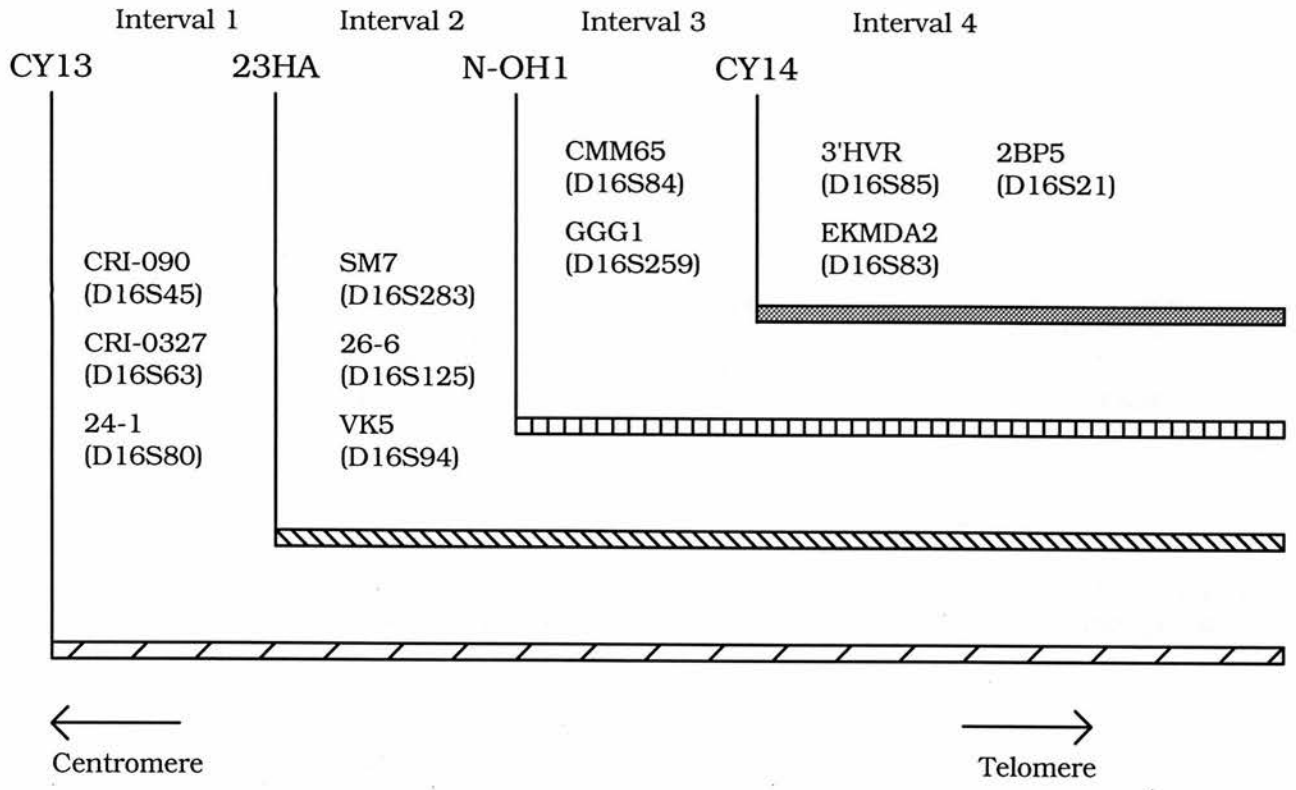
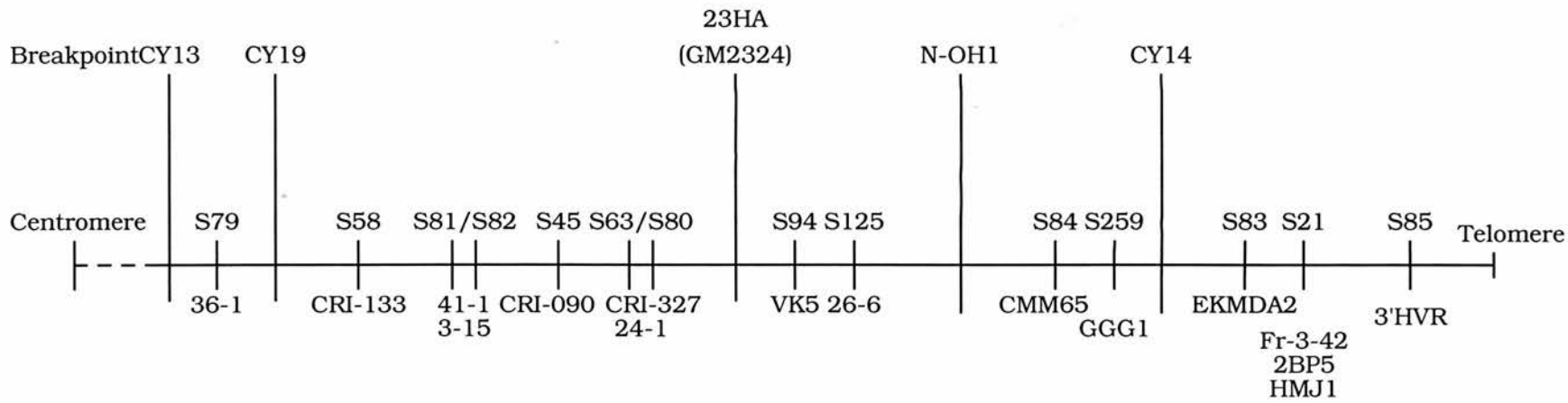


Figure 3.1 shows the order of the various breakpoints used in ordering markers located on chromosome 16p. The regions of 16p not present in each somatic cell hybrid are shaded (taken from Germino *et al.*, 1990).

Figure 3.2 DETAILED MAP OF CHROMOSOME 16p13



Map of the distal part of chromosome 16. The long vertical bars indicate breakpoints in chromosome 16 present in cell lines, the short vertical lines indicate polymorphic loci. The map is not to scale and is taken from Breuning *et al.* (1990a).

3.2.1 Linkage and heterogeneity analysis

Five-point linkage analyses were carried out using LINKMAP, with probe data from the 3'HVR (D16S85), CMM65 (D16S84), 26-6 (D16S125), and VK5 (D16S94) loci. The program was run using different fixed ratios of recombination in males to females, which identified a broad likelihood peak ratio in the region of 5 (θ_m:θ_f). This value was used in the final analysis. The resulting family lod scores were then analysed for linkage heterogeneity using the HOMOG program. This analysis was performed by P. Teague. Families found to be unlinked to chromosome 16p markers as a result of the heterogeneity analysis (PKD2 families) were then removed from the data set prior to further analysis.

3.2.2 Two-point linkage analysis

Two-point linkage analysis was performed between PKD1 and each of the five marker loci in turn, i.e. 3'HVR (D16S85), CMM65 (D16S84), SM7 (D16S283), 26-6 (D16S125), and VK5 (D16S94). Recombination fractions were estimated at intervals of 0.01 between 0 and 0.4. The MLINK subroutine was unable to support a fixed ratio of recombination between the two sexes, and thus sex-averaged lod scores were obtained. Two-point analysis was also performed for each of the pairs of marker loci in turn. MLINK was used to obtain sex-averaged lod scores in the first instance: analyses were then repeated using the ILINK subroutine, during which the male recombination fraction was set for each calculation, and the iterative process commenced at a female:male recombination ratio of 0.2, in order to identify changes in this ratio.

3.2.3 Multipoint linkage analyses

Multipoint analyses were run using LINKMAP as above, and probe data from the 3'HVR (D16S85), CMM65 (D16S84), SM7 (D16S283), 26-6 (D16S125), and VK5 (D16S94) loci. The program was unable to run if data from the multi-allelic loci 3'HVR (D16S85) and SM7 (D16S283) were entered at the same time, regardless of the total number of loci in the analysis. Therefore, a four-point analysis was run

Figure 3.3 MARKER LOCI USED IN LINKAGE DISEQUILIBRIUM STUDIES

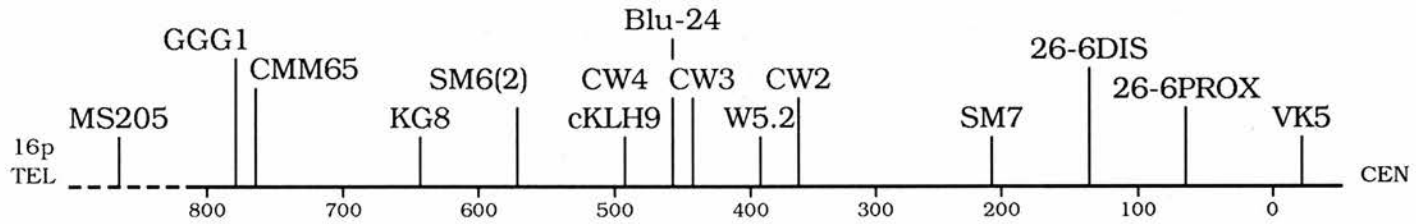


Figure 3.3 shows the approximate location of markers used by Dr. P. Harris and Ms. A. Snarey in the further analysis of recombinants and in a search for linkage disequilibrium. The positions of the markers are shown in relation to the CMM65 (D16S84), 26-6 (D16S125) and VK5 (D16S94) loci whch are thought to flank the PKD1 gene. Markers MS205, SM6(2), CW4 (D16S291), CW3, and CW2 were typed by Dr. P. Harris. Markers KG8, Blu-24, cKLH9 (D16S291), and W5.2 were typed by Ms. A. Snarey. The CA repeats CW4 and cKLH9 are identical.

using data from the PKD1, 3'HVR (D16S85), CMM65 (D16S84), and VK5 (D16S94) loci, and this was followed by a second run using data from the PKD1, CMM65 (D16S84), SM7 (D16S283), 26-6 (D16S125), and VK5 (D16S94) loci, in order to localise PKD1 more accurately.

3.2.4 Identification of recombinants

Haplotypes were drawn up for each pedigree, so as to allow the identification of recombinants. Once these had been identified, they were further characterised in collaboration with Dr. P. Harris (Oxford) and Ms. A. Snarey (London). Eight additional markers were used to try and identify the point at which recombination had taken place. Markers used were MS205, KG8, SM6 (2), CW4/cKLH9 (D16S291), Nik2.9, Blu-24, CW3, CW2, and CW1/W5.2 (Figure 3.3).

3.3 RESULTS

3.3.1 Linkage heterogeneity

The results of the heterogeneity analysis using HOMOG are shown in Table 3.1. The log (ln) likelihood, on the assumption of a single ADPKD locus (linkage homogeneity), was 55.22 compared with 63.05 (lod score 27.38) on the assumption of one linked and one unlinked ADPKD locus (linkage heterogeneity). The likelihood ratio is 2,514.9, which is statistically significant (chi-square 15.66; $P < 0.001$). The maximum likelihood value of α, the proportion of families linked, was found to be 0.81 with approximate 95% confidence limits of 0.54-0.97, showing that at least 3% of families are unlinked. The recombination fractions separating ADPKD from the anchor locus 3'HVR (D16S85) were 0.049 (males) and 0.010 (females) under linkage heterogeneity. Linkage heterogeneity is therefore present in this sample of families.

Table 3.1 RESULTS OF HETEROGENEITY ANALYSES

	Linkage heterogeneity	Linkage homogeneity
α (proportion of families linked	0.81 (0.54 - 0.97)	1.00
θ_m	0.049 (0.025 - 0.059)	0.047
θ_f	0.010 (0.005 - 0.012)	0.009
Log (ln) likelihood	63.05	55.22
P value	< 0.0001 (chi-square = 15.66)	
Likelihood ratio	2,514.9	

Results of heterogeneity analyses in 35 ADPKD kindreds. Multipoint lod scores for ADPKD, D16S85, D16S84, D16S125, and D16S94 were analysed with the HOMOG program. The maximum likelihood values of the proportion of families linked, α, and of the recombination fractions, measured proximally from D16S85, are shown for males (θ_m) and females (θ_f) separately with approximate 95% confidence limits (α, θ,) in parentheses.

3.3.2 Families that appear unlinked to PKD1

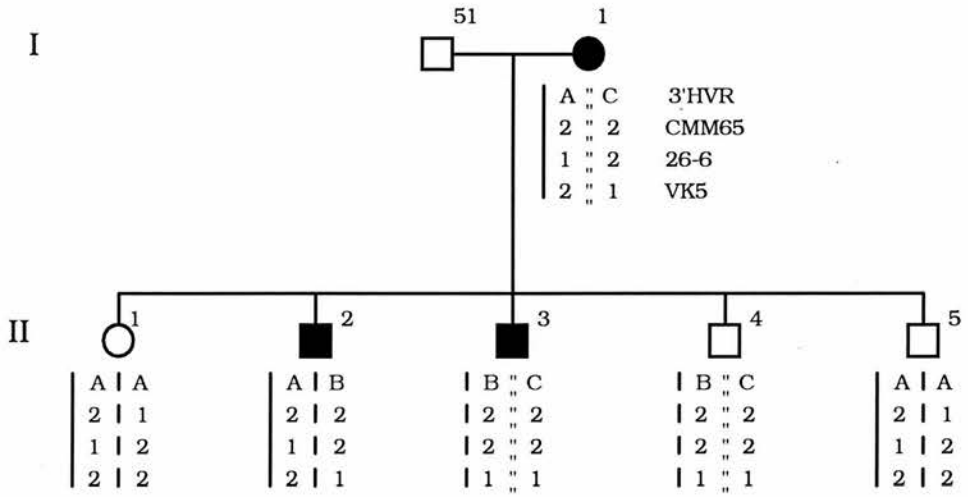
Only two small families out of 35 analysed show statistically significant evidence of non-linkage to PKD1 (Table 3.2), with probabilities of 0.013 (PK52) and 0.007 (PK53) that they are linked to PKD1, conditional on the maximum likelihood values of α (0.81) and θ_m (0.049). Family PK53 (Figure 3.4) is informative with flanking markers 3'HVR (D16S85) distally and 26-6 (D16S125) and VK5 (D16S94) proximally, and would require either two or three double recombinants out of five meioses to be consistent with a PKD1 locus. No exchange of flanking markers is seen. This family, therefore, does appear unlinked from haplotype analysis. Family PK52 (Figure 3.5a) is informative for 3'HVR (D16S85) and CMM65 (D16S84) distally but with none of the proximal markers, so that only a single recombinant (with 3'HVR and CMM65 in II-2) out of three meioses needs to be invoked to be consistent with a PKD1 locus. Further loci have been screened by P. Harris (Oxford)

Table 3.2 PROBABILITIES OF LINKAGE TO PKD1

Family number	Conditional probability of linkage to PKD1	Family number	Conditional probability of linkage to PKD1
PK 1	0.784	PK 30	0.997
PK 2	1.000	PK 31	0.795
PK 3	0.995	PK 32	0.967
PK 4	1.000	PK 35	0.895
PK 5	0.932	PK 36	0.930
PK 6	0.999	PK 37	0.751
PK 7	1.000	PK 38	0.886
PK 8	0.747	PK39	0.780
PK 11	0.999	PK41	0.860
PK 17	0.999	PK 42	0.998
PK 18	0.999	PK 45	0.961
PK 19	0.871	PK 49	0.958
PK 24	0.881	PK 50	0.608
PK 25	0.460	**PK 52**	**0.013**
PK 26	0.895	**PK 53**	**0.007**
PK 27	0.574	PK 55	0.602
PK 28	0.982	PK 56	0.886
PK 29	0.289		

Conditional probabilities of linkage to PKD1 are shown for each of the 35 kindreds studied. The results were obtained using the HOMOG program.

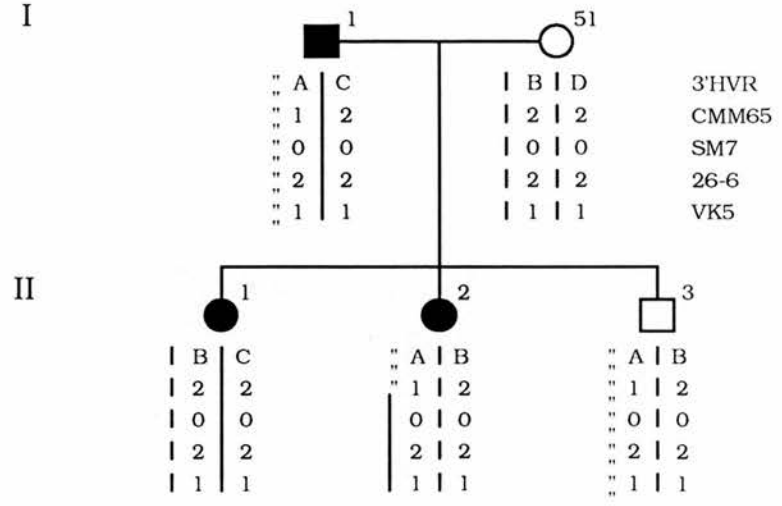
Figure 3.4 PEDIGREE OF FAMILY PK53



Haplotype analysis of family PK53 in which there are either two or three double recombinants with ADPKD, depending on the linkage phase, on the assumption of a PKD1 locus. There is no evidence of recombination between markers. The unaffected individuals II-1, II-4 and II-5 are aged 39, 32 and 31 years respectively. Possible haplotypes are shown for probes 3' HVR (D16S85), CMM65 (D16S84), 26-6 (D16S125) and VK5 (D16S94).

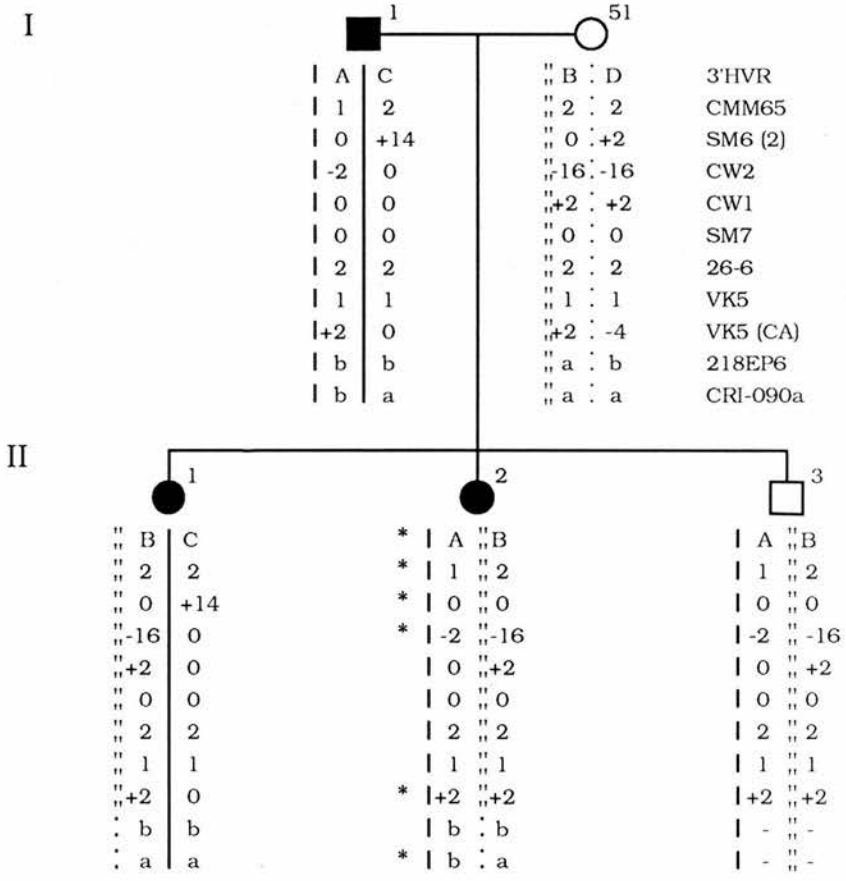
Figure 3.5 PEDIGREE OF FAMILY PK52

a)



Haplotype analysis of family PK52 in which a single recombination (in II-2) would be consistent with a PKD1 locus, although no exchange of flanking markers is demonstrated.

b)



Extended haplotype analysis of family PK52 demonstrating no exchange of flanking markers 3'HVR (D16S85) and CRI-090 (D16S45). * = "recombinant" loci in II-2.

in an attempt to find an informative proximal marker (Figure 3.5b). CA repeat polymorphisms isolated from the SM6 (2), CW2, VK5 (D16S94) and CRI-090a (D16S45) loci are informative, all of which were also recombinant in II-2, although there is no exchange of flanking markers (e.g. 3'HVR (D16S85) and CRI-090a (D16S45)). This would be consistent with non-linkage in this family also, since the alternative explanations of either a double crossover between CW2 and VK5 (D16S94), which are located approximately 400 kb apart, or a location for PKD1 proximal to CRI-090 (D16S45) are both very unlikely. These two families were therefore excluded from further analysis, and two-point and multipoint analyses were performed on the remaining 33 families.

3.3.3 Two-point linkage analysis

The results of two-point linkage analyses between PKD1 and the marker loci are shown in Table 3.3. Analyses were performed using MLINK so that sex-averaged values were obtained. The maximum likelihood value of the recombination fraction was found to be 0.07 at a lod score of 20.59 for PKD1-3'HVR (D16S85). The corresponding values of the recombination fractions were 0.03 at a lod score of 8.37 for PKD1-CMM65 (D16S84), 0.03 at a lod score of 15.85 for PKD1-SM7 (D16S283), 0.00 at a lod score of 5.26 for PKD1-26-6 (D16S125), and 0.00 at a lod score of 8.67 for PKD1-VK5 (D16S94). All loci show close linkage to PKD1 and highly significant lod scores. No recombination was found with the 26-6 (D16S125), and VK5 (D16S94) loci using MLINK, although three recombinants with 26-6 (D16S125) were subsequently identified by haplotype analysis. A single definite recombinant with CMM65 (D16S84) and a recombinant with SM7 (D16S283), which was later shown to be in an unlinked family, were identified after haplotype analysis (see below).

The results of two-point analyses between pairs of marker loci are shown in Table 3.4. In general, highly significant lod scores are obtained with small values of θ, confirming close linkage of these markers with each other. Relatively large values are obtained for θ_{max} between 3'HVR (D16S85) and 26-6 (D16S125) (0.16), SM7

Table 3.3 TWO-POINT LOD SCORES BETWEEN PKD1 AND MARKER LOCI (MLINK)

θ Locus	0.00	0.01	0.05	0.10	0.20	0.30	0.40	θ_{max}	Z_{max}
D16S85 (3'HVR)	-∞	15.168	20.297	20.230	16.119	10.115	3.787	0.07	20.585
D16S84 (CMM65)	-∞	7.947	8.280	7.535	5.348	2.915	0.817	0.03	8.366
D16S283 (SM7)	10.811	15.267	15.645	14.257	10.256	5.809	1.817	0.03	15.848
D16S125 (26-6)	5.259	5.171	4.748	4.119	2.813	1.632	0.603	0.00	5.259
D16S94 (VK5)	8.670	8.463	7.639	6.587	4.393	2.265	0.636	0.00	8.670

Results of two-point linkage analyses showing sex-averaged recombination fractions and the corresponding lod scores. The maximum likelihood values of the recombination fraction (θ_{max}) and lod scores (Z_{max}) are indicated.

Table 3.4 TWO-POINT LOD SCORES BETWEEN MARKER LOCI (MLINK)

θ Loci	0.00	0.01	0.05	0.10	0.20	0.30	0.40	θ_{max}	Z_{max}
D16S85 - D16S84	- ∞	6.986	9.454	9.328	7.052	3.904	1.13	0.07	9.566
D16S85 - D16S283	- ∞	13.797	18.686	18.673	14.58	8.621	2.768	0.07	18.996
D16S85 - D16S125	- ∞	- 11.902	- 0.33	2.953	3.524	2.215	0.746	0.16	3.732
D16S85 - D16S94	- ∞	- 1.183	7.15	9.068	7.867	4.631	1.369	0.11	9.145
D16S84 - D16S283	- ∞	4.616	5.236	4.847	3.432	1.85	0.54	0.04	5.238
D16S84 - D16S125	- ∞	3.343	4.094	3.825	2.63	1.32	0.316	0.05	4.094
D16S84 - D16S94	- ∞	0.63	1.195	1.247	0.946	0.523	0.155	0.08	1.257
D16S283 - D16S125	- ∞	- 6.876	- 0.659	1.174	1.642	0.991	0.276	0.17	1.701
D16S283 - D16S94	- ∞	- 0.17	3.301	3.971	3.291	1.879	0.56	0.11	3.979
D16S125 - D16S94	- ∞	4.604	5.158	4.649	3.003	1.379	0.328	0.04	5.18

Results of two-point linkage analyses between marker loci showing sex-averaged recombination fractions and the corresponding lod scores. The maximum likelihood values of the recombination fraction (θ_{max}) and lod scores (Z_{max}) are indicated. Values were obtained using the MLINK program.

Table 3.5 TWO-POINT LOD SCORES BETWEEN MARKER LOCI (ILINK)

θm Loci	0.00	0.01	0.05	0.10	0.20	0.30	0.40	θmax (males)	$Z max$
D16S85 - D16S84	- ∞	7.64	9.79	5.26	4.70	3.47	2.09	0.05	9.79
D16S85 - D16S283	- ∞	14.57	13.33	12.62	12.04	8.95	4.66	0.04	18.46
D16S85 - D16S125	- ∞	2.55	3.86	4.13	3.70	2.78	1.84	0.11	4.14
D16S85 - D16S94	- ∞	6.90	8.81	9.20	8.54	7.04	5.16	0.11	9.21
D16S84 - D16S283	- ∞	6.27	4.58	4.02	2.95	1.91	1.01	0.01	6.27
D16S84 - D16S125	- ∞	3.38	3.24	2.70	1.70	0.81	0.20	0.03	3.42
D16S84 - D16S94	- ∞	1.10	1.48	1.37	1.15	0.83	0.52	0.04	1.50
D16S283 - D16S125	- ∞	-	3.21	2.87	1.94	1.02	0.36	0.04	3.22
D16S283 - D16S94	- ∞	2.44	3.94	4.11	3.58	2.04	1.23	0.09	4.12
D16S125 - D16S94	- ∞	5.13	5.24	2.51	2.14	1.39	0.56	0.05	5.24

Results of two-point linkage analyses between marker loci showing male recombination fractions and the corresponding lod scores. The maximum likelihood values of the recombination fraction (θ_{max}) and lod scores (Z_{max}) are indicated. Values were obtained using the ILINK program and allowing the set ratio for recombination to vary.

(D16S283) and 26-6 (D16S125) (0.17), and SM7 (D16S283) and VK5 (D16S94) (0.11). In these cases, the lod scores are low, indicative of the small number of informative meioses.

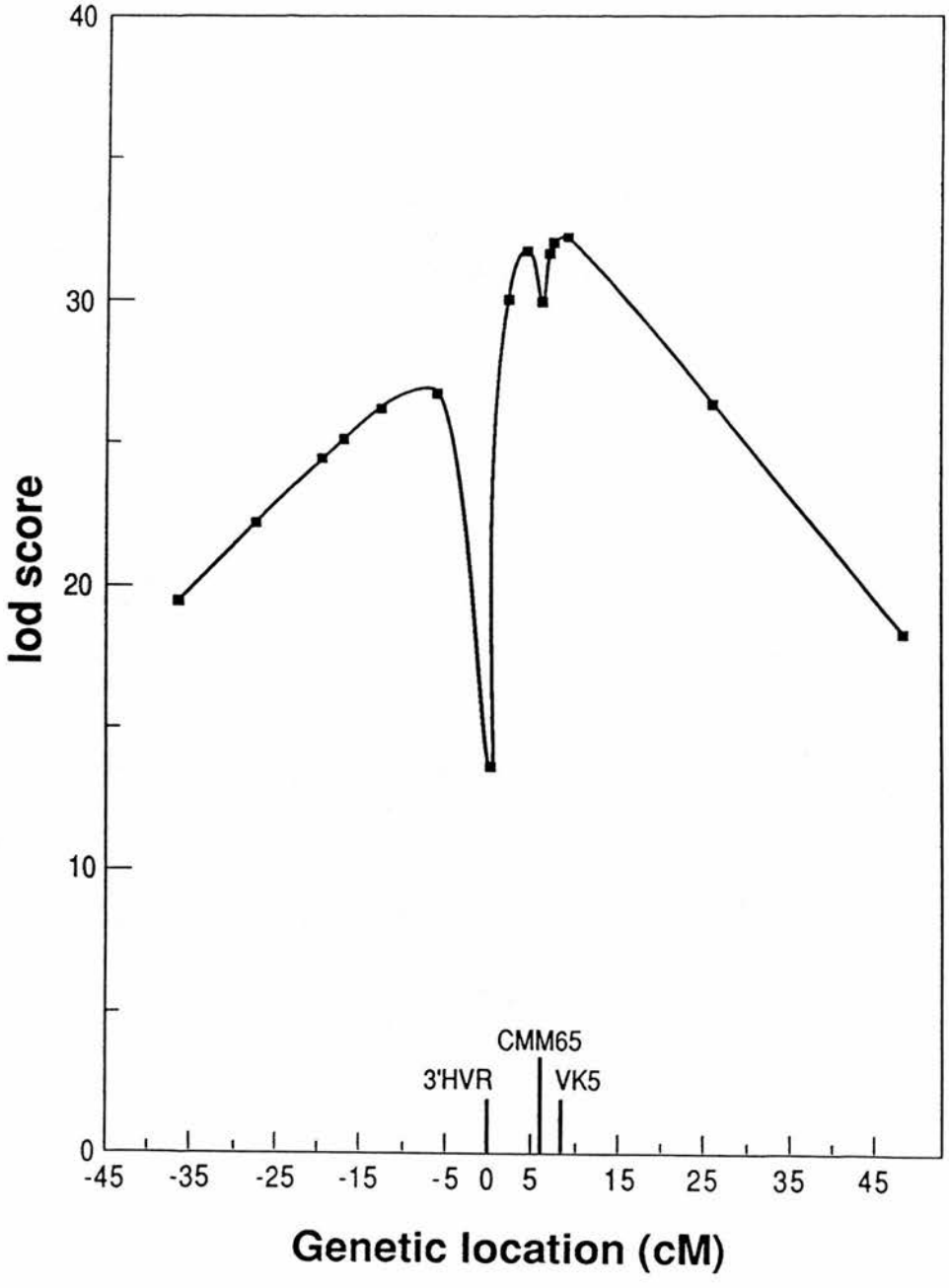
The two-point analyses between markers were repeated using ILINK with the iterative process commenced at various fixed values of θ_m (0.01 - 0.4), and a female: male recombination ratio of 0.2 (not fixed). The results are shown in Table 3.5, with the lod scores given for the corresponding θ_m. The values obtained for the male recombination fractions between markers correlate more closely with each other than the sex-averaged values of θ obtained with MLINK. However, during the course of these calculations, surprising values for the $\theta_f : \theta_m$ ratios were obtained and this was therefore studied in further detail (see below).

3.3.4 Multipoint linkage analysis

The results of the four-point linkage analysis with PKD1, 3'HVR (D16S85), CMM65 (D16S84), and VK5 (D16S94) are shown in Figure 3.6. The maximum likelihood is found to occur at the location D16S85-0.06-D16S84-0.02-(PKD1,D16S94)-cen (order 1) at a peak lod score of 32.16. Since no recombination was observed between PKD1 and D16S94, the order of these two loci cannot be determined. The likelihood of the order D16S85-0.04-PKD1-0.02-D16S84-0.02-D16S94-cen (order 2) is only 2.6 times lower than order 1, with PKD1 proximal to D16S84. However, order 1 is at least 2.7×10^5 times more likely than with PKD1 distal to D16S85. The order D16S85-0.06-D16S84-0.02-(PKD1,D16S94)-cen is also consistent with the recombinant haplotype in family PK19 (see below).

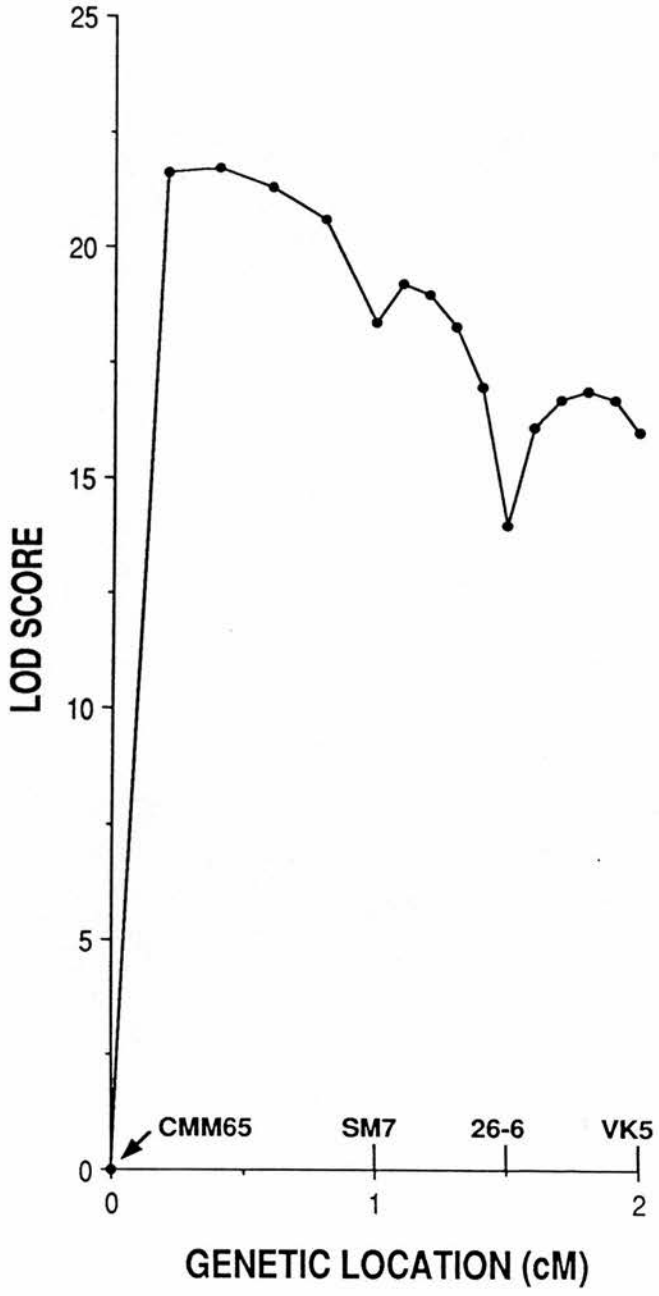
Given the most likely order, which places PKD1 proximal to CMM65 (D16S84), a five point linkage analysis was performed with PKD1, CMM65 (D16S84), 26-6 (D16S125), SM7 (D16S283), and VK5 (D16S94), so as to obtain a more accurate localisation of PKD1. The results of the five-point analysis are shown in Figure 3.7. The maximum likelihood is found to occur at the location D16S84-0.004-PKD1-0.006-D16S283-0.005-D16S125-0.005-D16S94-cen at a peak lod score of 21.64. The likelihood of this order is 279.25 times greater than the order

Figure 3.6 RESULTS OF THE FOUR-POINT LINKAGE ANALYSIS



Results of multipoint linkage analysis of ADPKD and the 3'HVR (D16S85), CMM65 (D16S84) and VK5 (D16S94) loci. The multipoint lod scores are plotted against genetic location (based on recombination fractions in males) with reference to the 3'HVR (D16S85) locus at position 0. The location of the CMM65 (D16S84) and VK5 (D16S94) loci are also shown. Distances are given in centiMorgans (cM).

Figure 3.7 RESULTS OF THE FIVE-POINT LINKAGE ANALYSIS



Results of multipoint analysis of ADPKD and the CMM65 (D16S84), 26-6 (D16S125), SM7 (D16S283) and VK5 (D16S94) loci. The multipoint lod scores are plotted against genetic location based on the recombination fraction in males.

D16S84-0.01-D16S283-0.001-PKD1-0.004-D16S125-0.005-D16S94-cen. The odds ratios for the likelihood of the possible locations of PKD1 are given in Table 3.6.

Table 3.6 ODDS RATIO FOR THE LOCALISATION OF PKD1 BY MULTIPOINT ANALYSIS

Order of Loci	Lod Score (Z_{max})	Odds Ratio
D16S84-0.004-PKD1-0.006-D16S283-0.005-D16S125-0.005-D16S94-cen	21.636	1
D16S84-0.01-D16S283-0.001-PKD1-0.004-D16S125-0.005-D16S94-cen	19.19	279.25
D16S84-0.01-D16S283-0.005-D16S125-0.003-PKD1-0.002-D16S94-cen	16.853	60,673.63

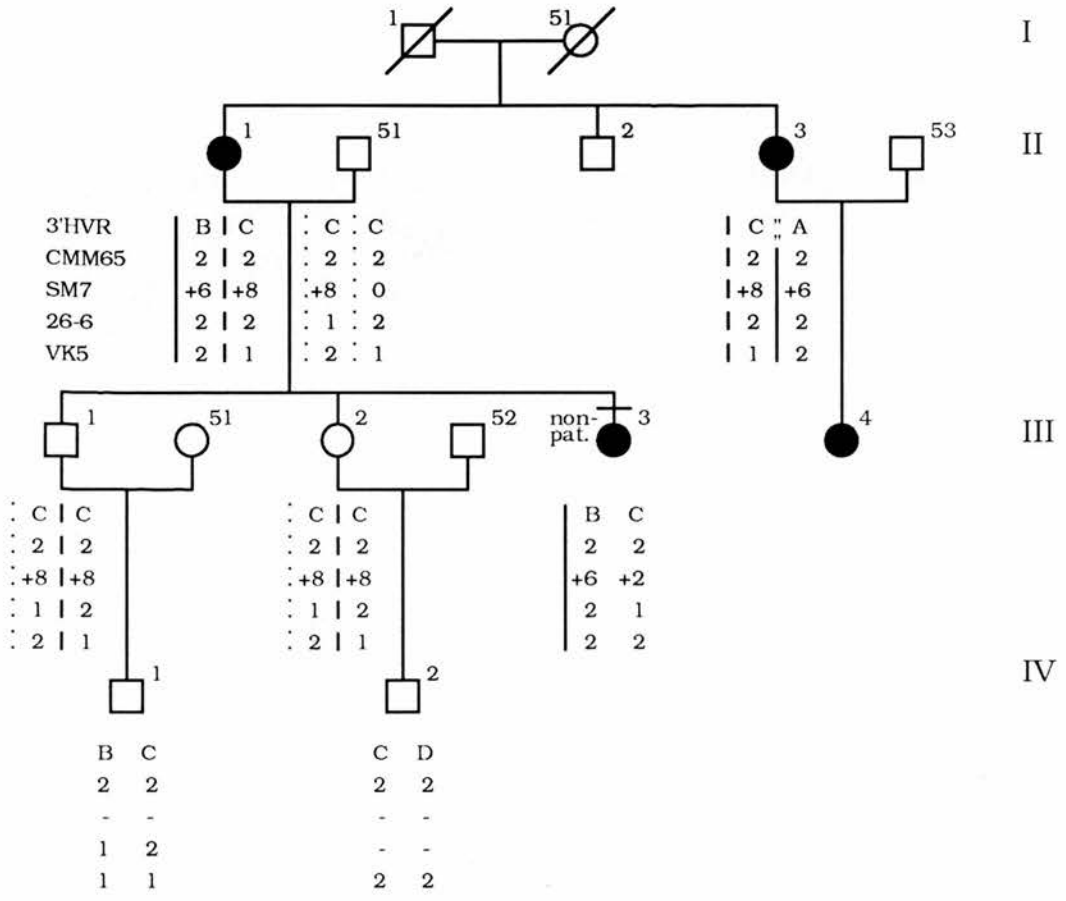
Table showing the maximum lod scores obtained during five-point analysis by LINKMAP of the CMM65 (D16S84), SM7 (D16S283), 26-6 (D16S125), VK5 (D16S94) and PKD1 loci. The odds ratio for each of the three possible orders has been calculated.

3.3.5 Analysis of PKD1 recombinants

Fourteen individuals, from eleven different kindreds, were identified as having a single recombination event between PKD1 and flanking markers.

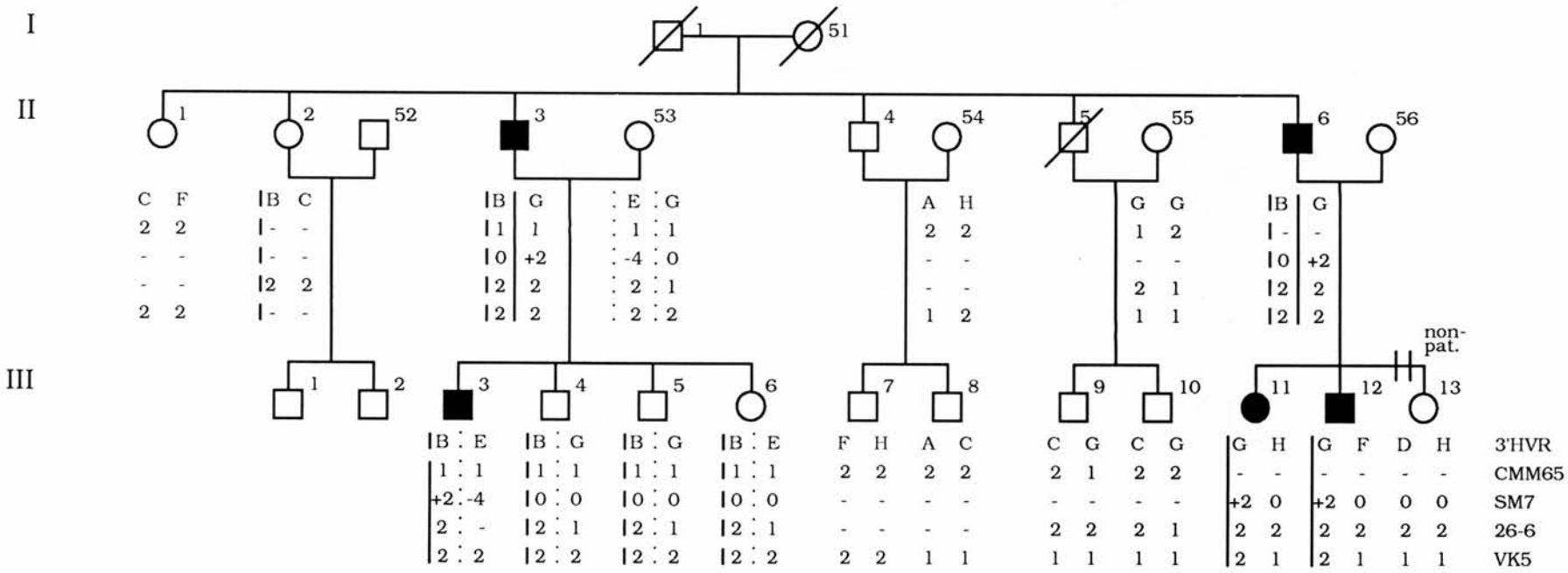
In family PK1 (Figure 3.8) individual II-3 is affected and recombinant for 3'HVR (D16S85), presumed uninformative for CMM65 (D16S84) and non-recombinant for SM7 (D16S283). Other recombinants for 3'HVR (D16S85), which are also uninformative for CMM65 (D16S84) and non-recombinant for SM7 (D16S283), are seen in family PK3 (Figure 3.9, individual III-3), family PK11 (Figure 3.10, individual V-9), family PK25 (Figure 3.11, individual III-2) and family PK31 (Figure 3.12, individual II-5). All these individuals are affected with a clearly positive ultrasound scan. Since the probe CMM65 (D16S84) is uninformative in all these recombinants, probe MS205 (Figure 3.3) was used to type the families

Figure 3.8 PEDIGREE OF FAMILY PK1



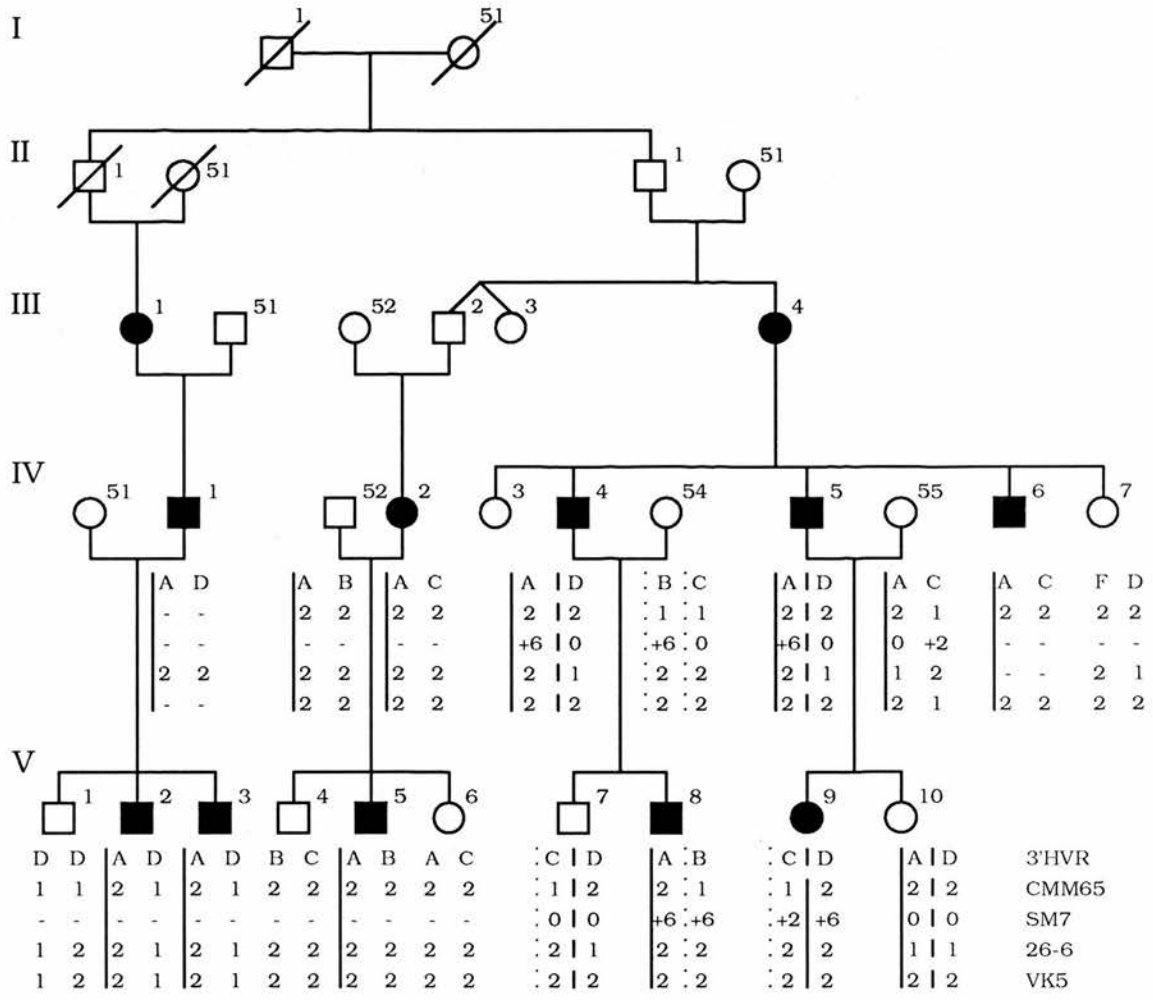
Haplotype analysis in family PK1 showing a single recombinant with 3'HVR (D16S85) (II-3), which is presumed uninformative for CMM65 (D16S84) and non-recombinant for SM7 (D16S283). Individual III-3 exhibits non-paternity on haplotype analysis with at least one marker (SM7), or new mutation at the SM7 locus. Further analysis with other markers (P. Harris) confirms non-paternity.

Figure 3.9 PEDIGREE OF FAMILY PK3



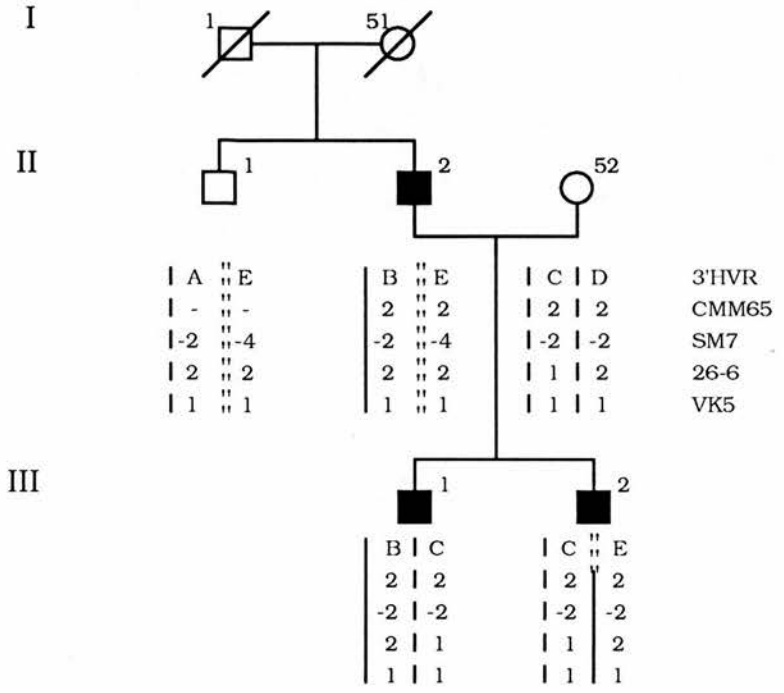
Haplotype analysis of family PK3 showing a recombination between the PKD1 gene and 3' HVR (D16S85) in individual III-3. This individual is uninformative with CMM65 (D16S84) and non-recombinant with SM7 (D16S283). III-13 exhibits non-paternity with at least two markers (3'HVR and VK5).

Figure 3.10 PEDIGREE OF FAMILY PK11



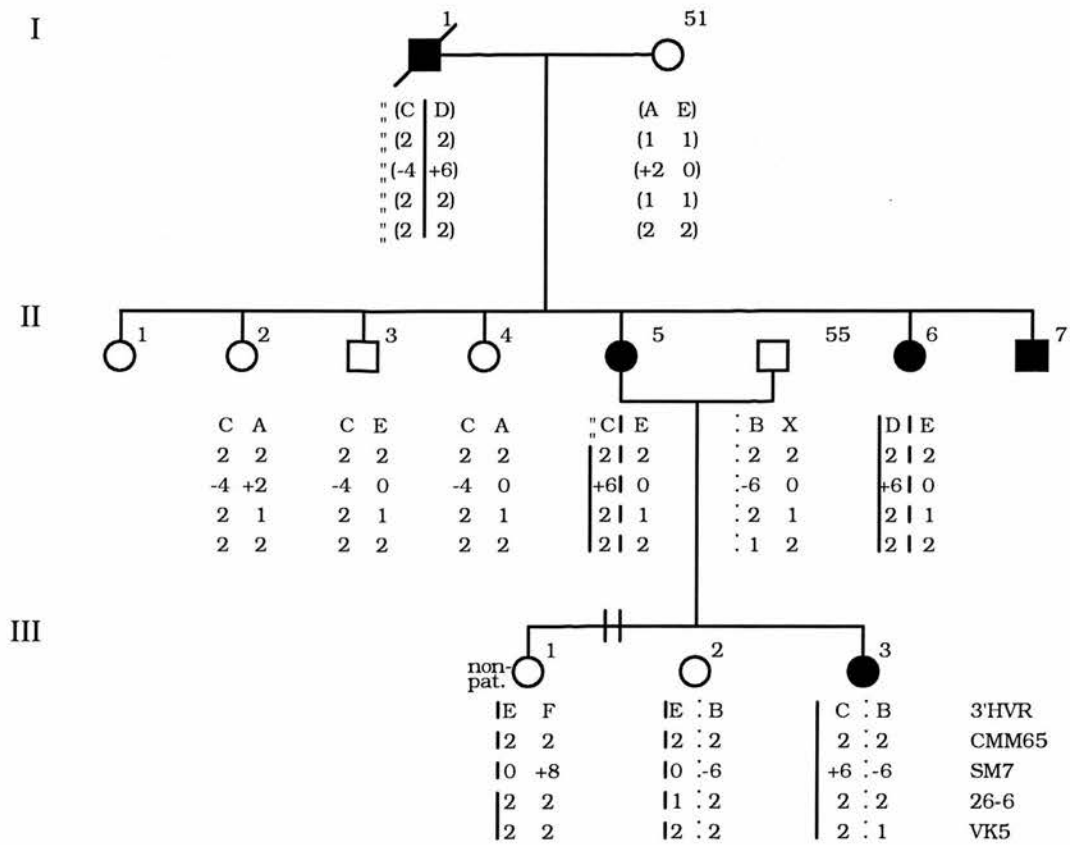
Haplotype analysis of family PK11 showing recombination in individual V-9. She is recombinant for 3'HVR (D16S85), uninformative with CMM65 (D16S84) and VK5 (D16S94), and non-recombinant with SM7 (D16S283) and 26-6 (D16S125).

Figure 3.11 PEDIGREE OF FAMILY PK25



Haplotype analysis of family PK25 showing a recombinant with 3'HVR (D16S85), which is uninformative with CMM65 (D16S84), 26-6 (D16S125) and VK5 (D16S94), and non-recombinant with SM7 (D16S283) (individual III-2).

Figure 3.12 PEDIGREE OF FAMILY PK31



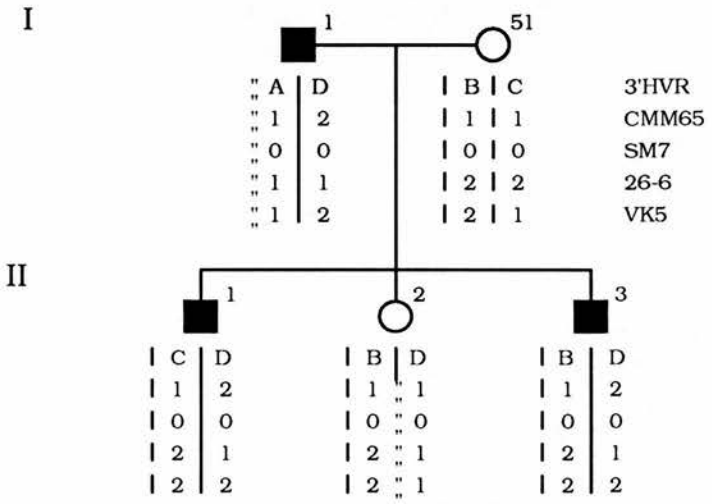
Haplotype analysis of family PK31, showing a recombinant with 3'HVR (D16S85) in individual II-5, which is then inherited by III-3. The typings of I-1 and I-51 are inferred by haplotype as no DNA was available for analysis. II-5 is uninformative with CMM65 (D16S84) and non-recombinant with SM7 (D16S283). Further typing (see text) shows the crossover to be between 3'HVR (D16S85) and MS205, which is located distal to CMM65 (D16S84), and therefore PKD1 is proximal to 3'HVR (D16S85). Individual III-1 exhibits genetic evidence of non-paternity with at least two markers. III-1 also shows a crossover in which PKD1 is non-recombinant with SM7 (D16S283) but recombinant with 26-6 (D16S125). III-1 is unaffected on ultrasound scan at the age of 31 years.

(P. Harris, Oxford). MS205 detects an RFLP and maps close to EKMDA (D16S83) which is proximal to CMM65 (D16S84). It was informative in all of the above recombinants. All were non-recombinant for MS205, which localised the crossovers to between 3'HVR (D16S85) and MS205, and the PKD1 gene to a position proximal to 3'HVR (D16S85).

In family PK5 (Figure 3.13) II-2 is recombinant for 3'HVR (D16S85) and non-recombinant for CMM65 (D16S84) and VK5 (D16S94). This individual is unaffected with a negative ultrasound scan at the age of 37 years, and is therefore unlikely to have a false negative diagnosis. In family PK7 (Figure 3.14) III-3 is also recombinant for 3'HVR (D16S85) and, by inference from the haplotype, non-recombinant for CMM65 (D16S84) and proximal markers. III-3 is affected with ADPKD. In family PK28 (Figure 3.15) individual II-2 is recombinant for 3'HVR (D16S85) and non-recombinant for CMM65 (D16S84) and other proximal markers. In these three families, the crossovers have occurred between 3'HVR (D16S85) and CMM65 (D16S84), again placing the disease gene proximal to 3'HVR (D16S85).

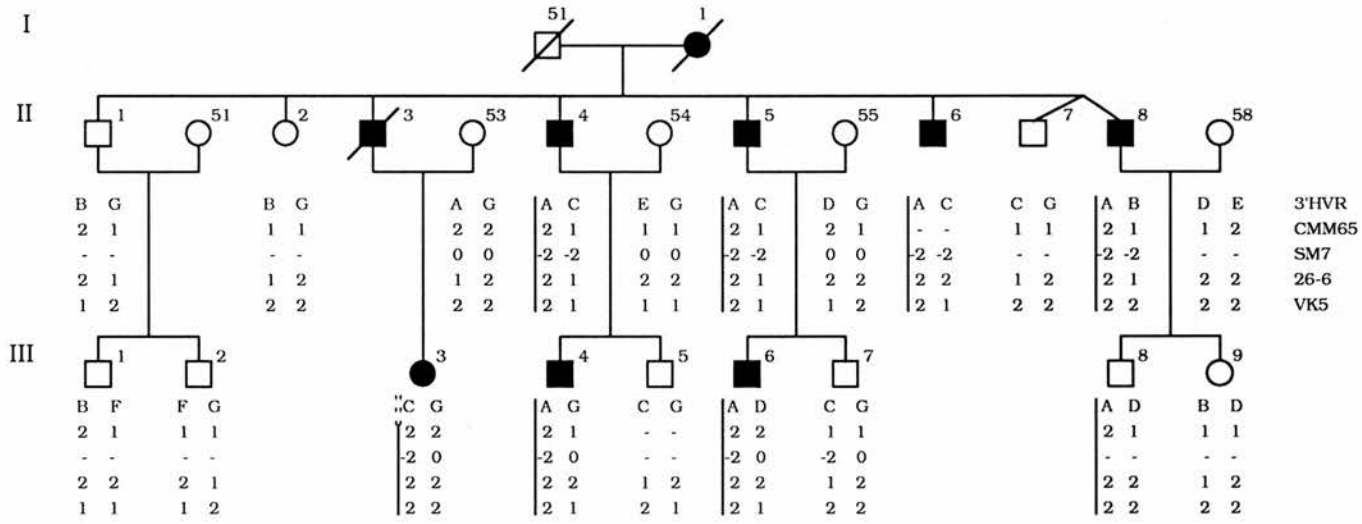
In family PK27 (Figure 3.16), three recombinants have occurred, all in affected individuals. II-4 is recombinant with 3'HVR (D16S85), uninformative with SM7 (D16S283), and non-recombinant with 26-6 (D16S125). II-1 and II-3 are both non-recombinant with 3'HVR (D16S85), uninformative with SM7 (D16S283), and recombinant with 26-6 (D16S125). Typing with extra probes (Figure 3.3) (A. Snarey and P. Harris) has helped in establishing the points of crossover. II-4 is non-recombinant with KG8, which places the crossover between KG8 and 3'HVR (D16S85) and the PKD1 locus proximal to 3'HVR (D16S85). II-1 and II-3 are both non-recombinant with KG8, CW4/cKLH9 (D16S291) and CW2, uninformative for SM6(2), CW3 and CW1, and recombinant for 26-6 (D16S125). This places both crossovers between CW2 and 26-6 (D16S125), and the PKD1 locus distal to 26-6 (D16S125). The 26-6 (D16S125) typings were repeated and found to be consistent. Confirmation of these recombinants with other proximal markers was not possible, since I-1 was uninformative for VK5. The probability that this family is linked to PKD1 is only 0.57 based on the analysis shown in Table 3.2. However, exchange of

Figure 3.13 PEDIGREE OF FAMILY PK5



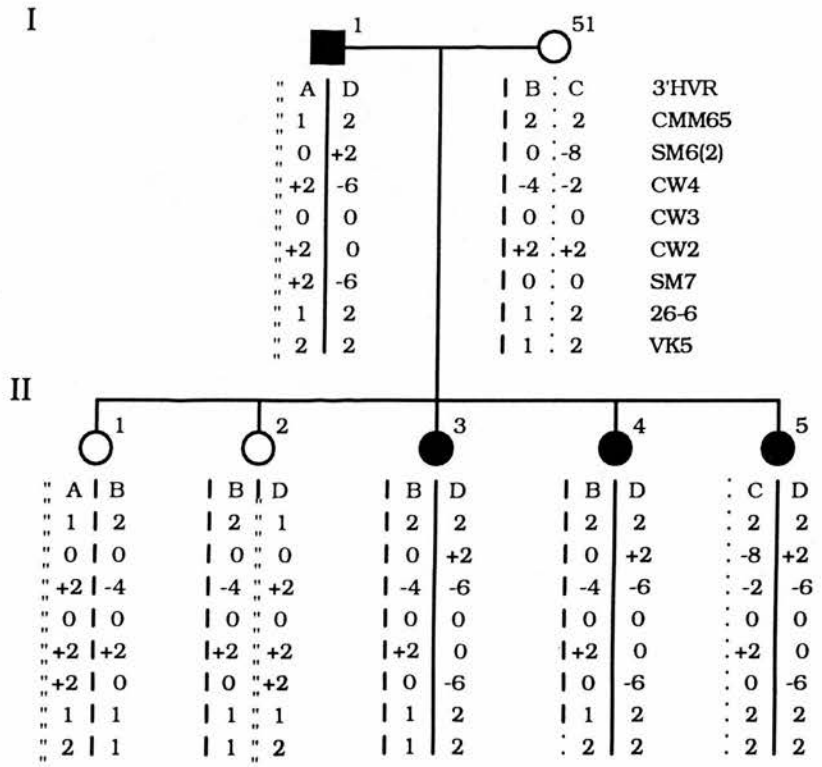
Haplotype analysis of family PK5 in which individual II-2 is recombinant for 3'HVR (D16S85) and non-recombinant for CMM65 (D16S84), placing PKD1 proximal to 3'HVR (D16S85). II-2 is unaffected with a negative ultrasound scan at age 37 years.

Figure 3.14 PEDIGREE OF FAMILY PK7



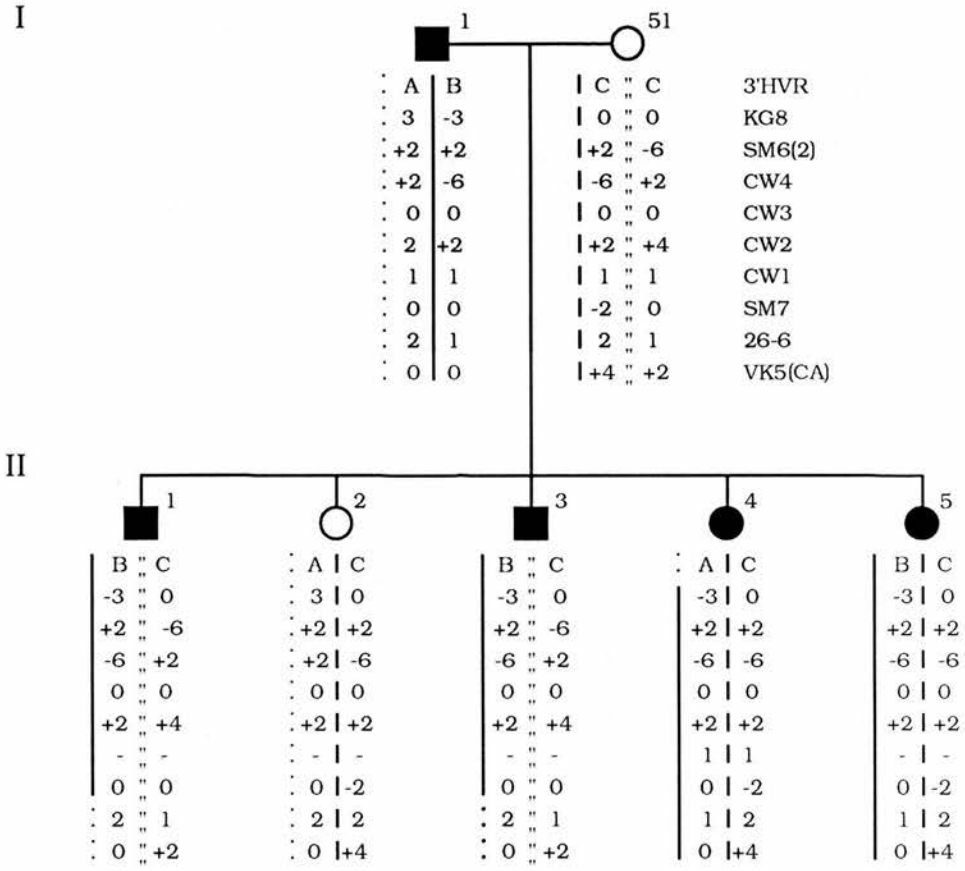
Haplotype analysis of family PK7, with the disease haplotype marked. Individual III-3 is apparently recombinant with 3'HVR (D16S85), but since II-3 could not be typed, it is not possible to establish that marker exchange has occurred. III-8 has received the disease haplotype, but is currently unaffected at age 19 years.

Figure 3.15 PEDIGREE OF FAMILY PK28



Haplotype analysis of family PK28 showing a recombinant with 3'HVR (D16S85) in individual II-2. This individual is unaffected with a negative ultrasound scan at age 33 years.

Figure 3.16 PEDIGREE OF FAMILY PK27



Haplotype analysis of family PK27 showing three recombinants, all in affected individuals, II-1, II-3 and II-4. II-1 and II-3 are both non-recombinant with 3'HVR (D16S85), KG8, CW4 (D16S291) and CW2, and recombinant with 26-6 (D16S125). The other markers are uninformative. II-4 is recombinant with 3'HVR (D16S85) and non-recombinant with KG8, CW4 (D16S291), CW2 and 26-6 (D16S125). The other markers are uninformative.

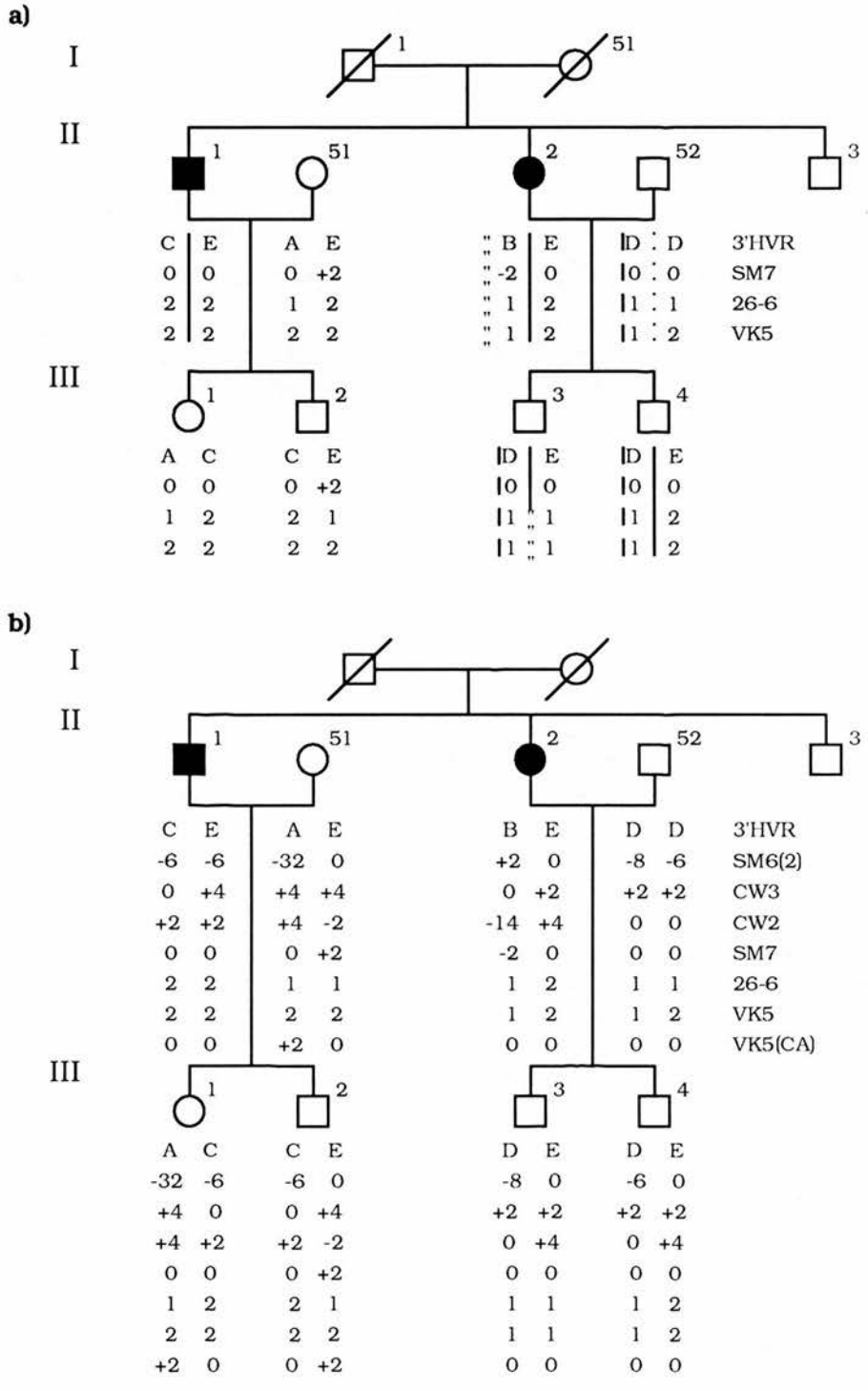
flanking markers has been shown to occur in all three recombinant meioses, suggesting that it may be linked to PKD1. Individual III-1 in family PK31 (Figure 3.12) exhibits genetic evidence of non-paternity with at least two markers. III-1 also shows a crossover in which PKD1 is non-recombinant with 3'HVR (D16S85), and SM7 (D16S283) but recombinant with 26-6 (D16S125). Markers CMM65 (D16S84) and VK5 (D16S94) were uninformative. The 26-6 (D16S125) typings were repeated and again found to be consistent; no informative proximal markers were available to confirm that recombination has occurred. This individual is unaffected at age 31 years with a negative ultrasound scan, and a false negative diagnosis is therefore unlikely. This recombinant again places the location of PKD1 distal to 26-6 (D16S125).

Individual III-3 in family PK29 (Figure 3.17a) was found to be recombinant with 3'HVR (D16S85) and SM7 (D16S283), and non-recombinant with 26-6 (D16S125) and VK5 (D16S94) after initial typing with these four markers. III-3 is unaffected at age 42 years with a negative ultrasound scan. This would have placed the location of PKD1 proximal to SM7 (D16S283). Further typing with other markers (P. Harris and A. Snarey), however, suggests that this family is unlinked to PKD1 (Figure 3.17b).

Individual III-5 in family PK19 (Figure 3.18) is recombinant for 3'HVR (D16S85) and CMM65 (D16S84), uninformative for SM7 (D16S283) and 26-6 (D16S125), and non-recombinant for VK5 (D16S94). Further typing (A. Snarey and P. Harris) shows III-5 to be uninformative for SM6(2) and CW4/cKLH9 (D16S291), but non-recombinant with CW3 and CW2. This places the crossover between CMM65 (D16S84) and CW3, and PKD1 proximal to CMM65 (D16S84). III-5 is, however, unaffected with a negative ultrasound scan at age 23 years, and therefore the possibility remains that he may develop the disease.

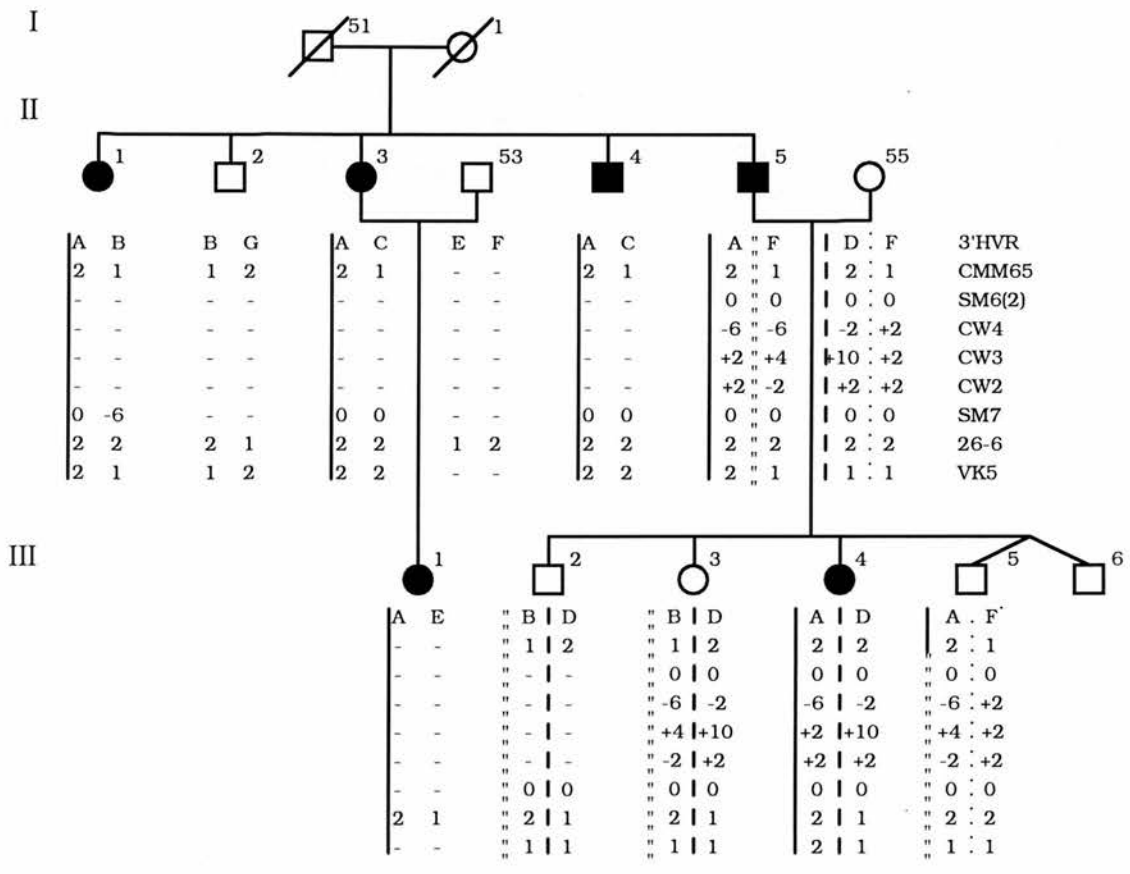
Taken together, the above results localise PKD1 distal to 26-6 (D16S125), and proximal to CMM65 (D16S84).

Figure 3.17 PEDIGREE OF FAMILY PK29



Haplotype analysis of family PK29, where initial typing with three markers suggested a recombinant with 3'HVR (D16S85) and SM7 (D16S283) in individual III-3 (Figure 3.17a). This individual was non-recombinant with 26-6 (D16S125) and VK5 (D16S94), which would have placed the location of PKD1 proximal to SM7 (D16S283). Further typing, however, suggests that this family is unlinked to PKD1 (Figure 3.17b). In individual III-3 recombination between SM7 (D16S283) and 26-6 (D16S125) has occurred, and in individual III-4 recombination between 26-6 (D16S125) and VK5 (D16S94) has occurred.

Figure 3.18 PEDIGREE OF FAMILY PK19



Haplotype analysis of family PK19 in which individual III-5 is recombinant for 3'HVR (D16S85) and CMM65 (D16S84), and non-recombinant with CW3, CW2 and VK5 (D16S94). The other markers are uninformative. III-5 is unaffected with a negative ultrasound scan at age 23 years.

3.3.6 Identification of new mutations

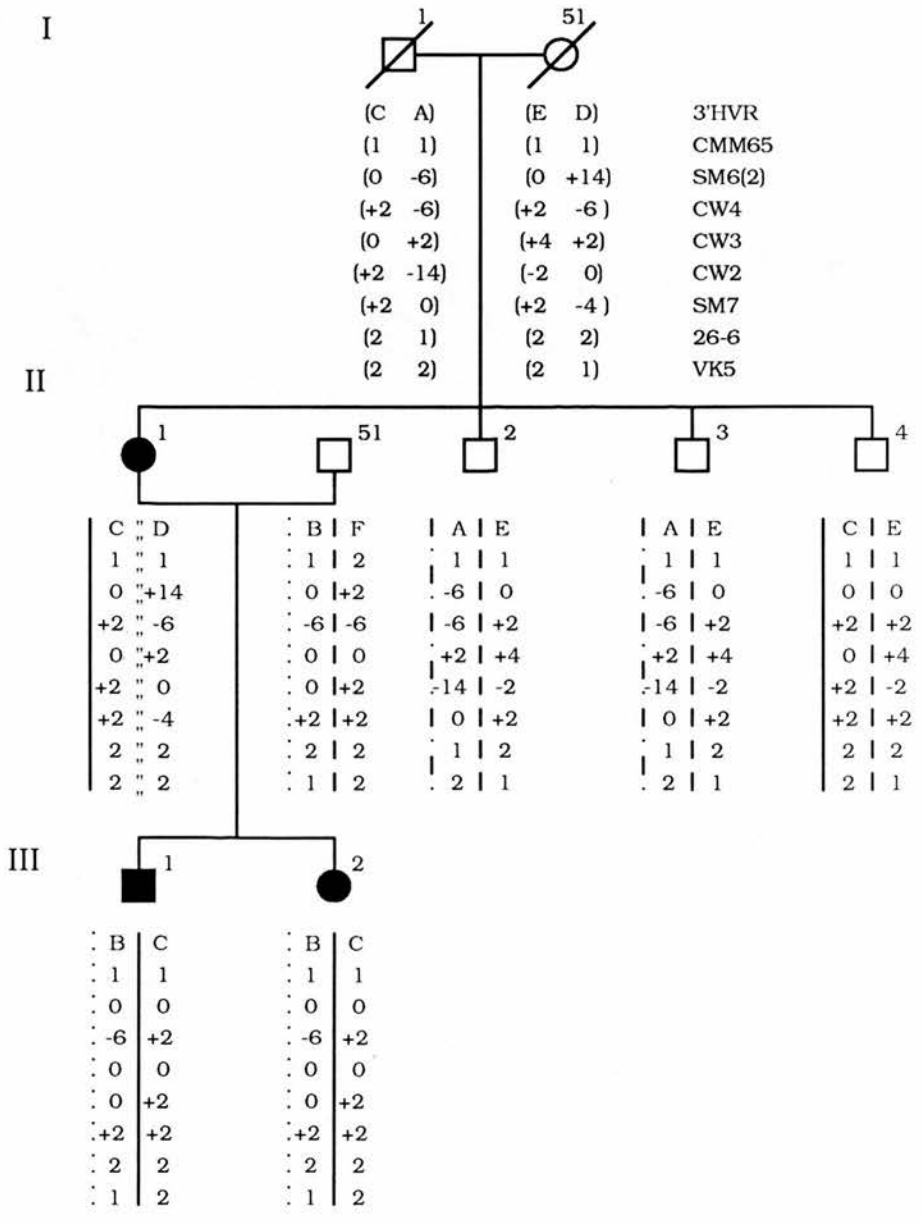
Following haplotype analysis, two individuals who had apparently inherited a new mutation were identified. These were in families PK8 and PK54. In family PK8 (Figure 3.19), both parents are unaffected as far as can be determined from death certificates. Individual II-1 appears to be a new mutation, with the affected chromosome being inherited by both III-1 and III-2. Individual II-4 has the same 'affected' chromosome, but is unaffected with a normal ultrasound scan at age 53 years. The possibility exists, however, that the family is unlinked to chromosome 16 (PKD2). Family PK54 (Figure 3.20) was excluded from linkage analysis in view of the fact that there was only one affected individual and no clear linkage to chromosome 16. I-1 is assumed to be unaffected, and I-51 is clearly so with a negative ultrasound scan at age 84 years. II-1 has now moved to Oxford where the two children have been typed by Dr. P. Harris. The children of II-1 are both affected and have inherited the same paternal chromosome from II-1. This chromosome originated from I-51 and has also been inherited by II-2 and II-6. Both of these siblings are unaffected, II-2 has a negative ultrasound scan at age 52 years.

3.3.7 Female : male recombination ratios in the PKD1 region

Twenty-nine single crossover events were identified in the region flanked by 3'HVR (D16S85) and VK5 (D16S94). Of these, ten occurred in female meioses, eighteen in male meioses, and one was unknown. Twelve recombinants occurred in the interval between 3'HVR (D16S85) and PKD1, and all of these were in male meioses. Nine recombinants occurred in the interval between PKD1 and VK5 (D16S94), three in male meioses and six in female meioses. Seven recombinants occurred in the overlapping region between CMM65 (D16S84) and 26-6 (D16S125), four in female meioses and three in male meioses. Although the total number of crossovers is small, there does seem to be a change from a clear excess of recombinants in males distally, to a female excess nearer VK5 (D16S94).

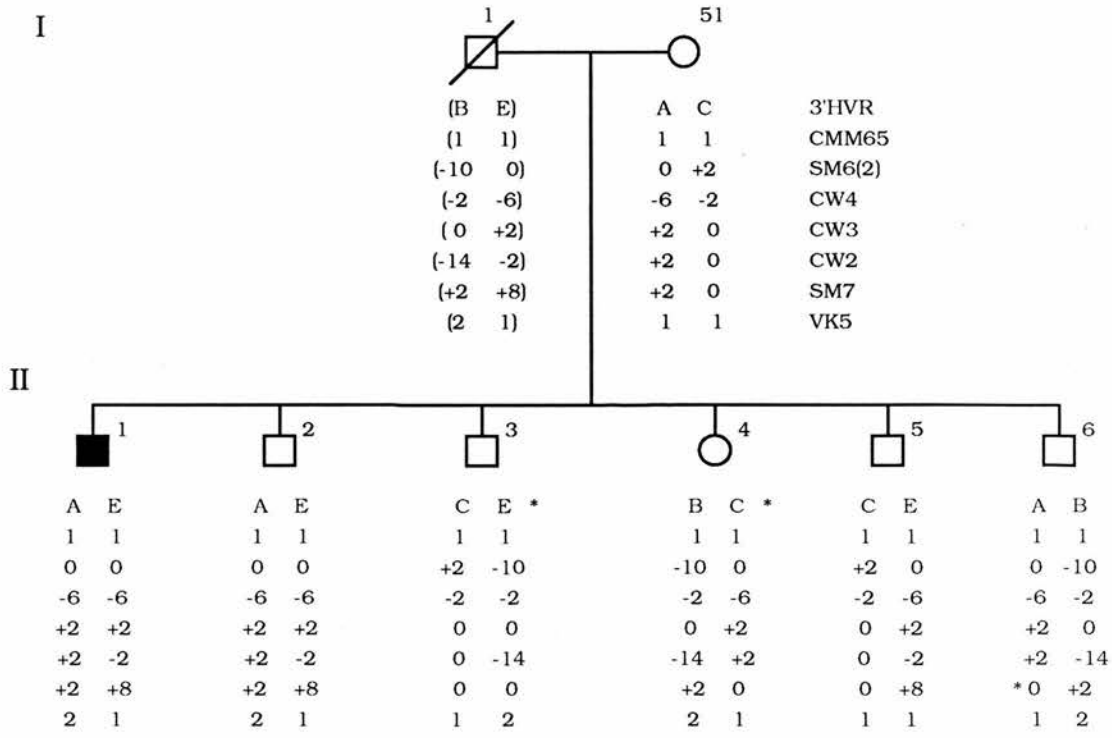
In the course of performing the two-point lod score analyses between markers using ILINK and a starting value of 0.2 for $\theta_f : \theta_m$ (female/male ratio),

Figure 3.19 PEDIGREE OF FAMILY PK8



Haplotype analysis of family PK8. The haplotypes of I-1 and I-51 are inferred. Individual II-1 appears to be a new mutation, with the affected chromosome being inherited by both III-1 and III-2. Individual II-4 has the same 'affected' chromosome, but without the mutation, and is unaffected with a normal ultrasound scan at age 53 years.

Figure 3.20 PEDIGREE OF FAMILY PK54



Haplotype analysis of family PK54 showing a new PKD1 mutation in II-1. I-1 is assumed to be unaffected having died at age 74 years of a cerebrovascular accident and prostatic carcinoma. His blood urea, electrolytes and creatinine were all normal. I-51 is unaffected with a normal ultrasound scan at age 84 years. II-1 has now married (family in Oxford; typed by Dr. P. Harris) and both his children have inherited the affected chromosome and are affected. The mutation has occurred on the maternal chromosome. Individuals II-3 and II-4 are recombinant for 3'HVR (D16S85), and individual II-6 is recombinant for SM7 (D16S283) and VK5 (D16S94).

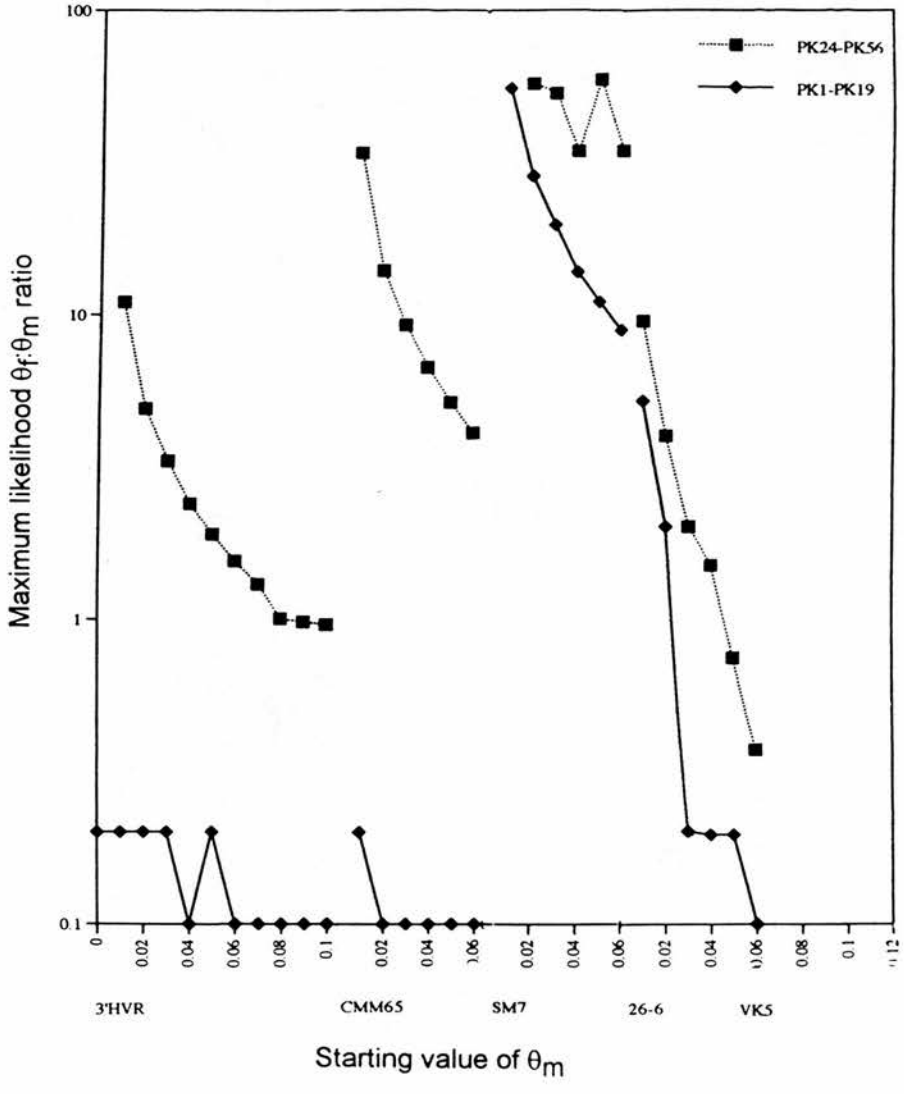
surprising values were obtained at the end of the iterative process. The analysis was performed in two runs, one for families PK1-PK19 and one for families PK24-PK56. Figure 3.21 is a graphical representation of the maximum likelihood $\theta_f : \theta_m$ ratio obtained at the various starting values of θ_m. For some of the pairs of markers [3'HVR (D16S85)-CMM65 (D16S84); CMM65 (D16S84)-SM7 (D16S283)] widely differing ratios were obtained in the two data sets, whereas for other pairs [SM7 (D16S283)-26-6 (D16S125); 26-6 (D16S125)-VK5 (D16S94)] the ratios are of the same order. This can at least partly be explained by the fact that in the distal region [3'HVR (D16S85)-SM7 (D16S283)] twice as many of the recombinants in males occurred in the first group of families (PK1-PK19) than in the second group (PK24-PK56), whereas in the proximal region [SM7 (D16S283)-VK5 (D16S94)]the recombinants are more evenly distributed between the two groups. In order to establish whether or not this was an artefact, further calculations were performed. Firstly, the ILINK program available was able to run with a maximum of 30 pedigrees, and therefore the first 30 pedigrees (PK1-PK45) were used. Runs were set up between pairs of markers, commencing the iterative process at the expected male recombination fraction for that interval. Each run was repeated using nine different set values, between 0.1 and 30.0, for the $\theta_f : \theta_m$ ratio, and the maximum likelihood value identified. The results are shown in Table 3.7.

Table 3.7 FEMALE : MALE RECOMBINATION RATIOS (ILINK)

Loci	Starting value of θ_m	θ_m at Z_{max}	θ_f at Z_{max}	$\theta_f : \theta_m$ ratio	Z_{max}
D16S85 - D16S84	0.06	0.074	0.016	0.2	10.97
D16S84 - D16S283	0.01	0.007	0.126	20.0	4.69
D16S283 - D16S125	0.005	0.042	0.465	30.0	2.84
D16S125 - D16S94	0.005	0.075	0.016	0.2	4.68

Results of two-point linkage analysis in families PK1-PK45 (n = 30) using ILINK showing male (θ_m) and female (θ_f) recombination fractions, and the ratio $\theta_f : \theta_m$, at the maximum lod score (Z_{max}) obtained during the run

Figure 3.21



Graphical representation of the maximum likelihood $\theta_f : \theta_m$ ratio obtained at various starting values of θ_m. The data were obtained using the ILINK program with families PK1-PK19 and then families PK24-PK56.

The female/male ratio, $\theta_f : \theta_m$, obtained at the maximum lod score for each marker-marker pair is 0.2 for the interval 3'HVR (D16S85)-CMM65 (D16S84), rising to 20.0 for CMM65 (D16S84)-SM7 (D16S283), 30.0 for SM7 (D16S283)-26-6 (D16S125), and falling again to 0.2 for the interval 26-6 (D16S125)-VK5 (D16S94).

Secondly, three-point analyses were run on all the families (PK1-PK56) using the LINKMAP program as follows: (1) CMM65 (D16S84) was used as a test locus at five positions between 3'HVR (D16S85) and SM7 (D16S283) which were fixed apart at a value of $\theta_m = 0.07$; (2) SM7 (D16S283) was used as a test locus at five locations between CMM65 (D16S84) and 26-6 (D16S125) which were fixed apart at a value of $\theta_m = 0.015$; (3) 26-6 (D16S125) was used as a test locus at five locations between SM7 (D16S283) and VK5 (D16S94) which were fixed apart at a value of $\theta_m = 0.01$; and (4) VK5 (D16S94) was used as a test locus at five locations for the outside interval with SM7 (D16S283) and 26-6 (D16S125) fixed apart at a value of 0.005. The results are shown in Table 3.8.

Table 3.8 FEMALE : MALE RECOMBINATION RATIOS (LINKMAP)

Locus order and θ_m	$\theta_f : \theta_m$ ratio	Z_{max}
D16S85-0.056-D16S84-0.016-D16S283	0.2	32.02
D16S84-0.003-D16S283-0.012-D16S125	35.0	10.72
D16S283-0.008-D16S125-0.002-D16S94	40.0	7.82
D16S283-0.005-D16S125-0.1-D16S94	2.0	7.31

Results of three-point analyses in all PKD1 families using LINKMAP showing the $\theta_f : \theta_m$ ratio obtained at Z_{max} for each of the four intervals studied

The female/male ($\theta_f : \theta_m$) ratio obtained for the interval 3'HVR (D16S85)-SM7 (D16S283) is 0.2 at a maximum lod score of 32.02, with CMM65 (D16S84) most likely to be located as follows:-

3'HVR (D16S85)-0.056-CMM65 (D16S84)-0.016-SM7 (D16S283).

Once more, the $\theta_f : \theta_m$ ratio rises markedly in the next two intervals being 35.0 at a maximum lod score of 10.72 for CMM65 (D16S84)-0.003-SM7 (D16S283)-0.012-26-6 (D16S125), and 40.0 at a maximum lod score of 7.82 for SM7 (D16S283)-0.008-26-6 (D16S125)-0.002-VK5 (D16S94). The $\theta_f : \theta_m$ ratio then falls again to 2.0 at a maximum lod score of 7.31 for SM7 (D16S283)-0.005-26-6 (D16S125)-0.1-VK5 (D16S94).

3.4 DISCUSSION

3.4.1 Genetic heterogeneity

In this study, 35 families with ADPKD have been analysed with five polymorphic markers on chromosome 16. The data were used to calculate α, the proportion of PKD1-linked families versus unlinked families with a mutation elsewhere in the genome. The likelihood of two loci is 2,514.9 times greater than for a single locus, with two small families providing statistical evidence of non-linkage to PKD1, confirmed on haplotype analysis. A third unlinked family, PK29, was identified only after extended haplotype analysis. The maximum likelihood value of α is 0.81, with 95% confidence limits of 0.54 - 0.97, showing that at least 3% of families are unlinked. Several attempts have been made to estimate α in different populations. At first, no evidence for heterogeneity was found in a set of 28 families from northern Europe, since the maximum likelihood value of α was 1.00 (Reeders *et al.*, 1987). All of the families studied by Lazarou *et al.* (1987) and Turco *et al.* (1991) demonstrated evidence of linkage to chromosome 16p. Mandich *et al.* (1990) found no significant evidence for genetic heterogeneity after examining twenty-nine Italian families, although the maximum likelihood value of α was 0.92. Two unlinked families of Italian origin had been described earlier (Kimberling *et al.*, 1988; Romeo *et al.*, 1988), so it seems that heterogeneity is present in this population, albeit at a

low level. A value of $\alpha = 0.95$ was obtained after analysis of 12 French families together with the 28 families analysed by Reeders *et al.* (1987), but this was not statistically different from $\alpha = 1.00$ (Bachner *et al.*, 1990). A similar value ($\alpha = 0.96$) was observed by Pieke *et al.* (1989), after analysis of 42 families from the USA, but again statistically significant heterogeneity was not present.

The value of $\alpha = 0.81$ obtained in this study is lower than the above estimates, but similar to a value of 0.86 obtained by Peters and Sandkuijl (1992) who analysed a large set of 328 families from all over Europe, including the 35 described here. As in the present study, the 95% confidence limits were wide (0.79 - 0.91), reflecting the relatively large number of small families where the lod score was close to zero. This heterogeneity analysis, and those of Reeders *et al.* (1987), Bachner *et al.* (1990), and Mandich *et al.* (1990), were performed on two-point linkage data, and Pieke *et al.* (1989) used three point data. In the present study, however, the more powerful technique of multipoint linkage analysis followed by heterogeneity analysis is employed.

Phenotypic differences between PKD1 and PKD2 families have been described (Bear *et al.*, 1989; Parfrey *et al.*, 1990; Ravine *et al.*, 1992) where PKD2 families have a milder phenotype with delayed appearance of cysts, and a lower risk of developing hypertension or progressing to renal failure. It is therefore possible that proportionately fewer PKD2 families attend renal clinics and are consequently under-represented in genetic studies. Such selection bias, if present, would under-estimate the true number of unlinked families. Alternatively, the milder phenotype may represent allelic differences confined to the limited number of PKD2 families analysed. The value of α obtained in this study of families ascertained at renal clinics may, if anything, under-estimate the proportion of PKD2 families. The phenotype in family PK53 (Figure 3.4) is consistent with a milder form of the disease, with the mother (I-1) having normal renal function and only borderline hypertension at age 59 years. However, this is not uniformly so within the family, since the two affected sons (II-2 and II-3) required renal dialysis at the ages of 33 years and 27 years respectively. Family PK52 (Figure 3.5) shows the typical PKD1

phenotype: the father (I-1) received a renal transplant at age 64 years, and the affected daughters (II-1 and II-2) are hypertensive with normal renal function at age 43 years and 39 years respectively. In family PK29 (Figure 3.17) II-1 required dialysis at the age of 51 years and was transplanted the following year. II-2 has been lost to follow up. All three families are of Scottish ancestry. The three unlinked families seen in this study are similar to those first described by Romeo *et al.* (1988) and Kimberling *et al.* (1988) in that they demonstrate the typical PKD1 phenotype.

This result has implications for pre-symptomatic diagnosis of ADPKD. The available evidence suggests that there is little indication for this under the age of 18 years (Watson *et al.*, 1990), but in early adulthood pre-symptomatic diagnosis has the advantage of allowing close follow-up of individuals at risk. The level of heterogeneity observed will have an effect on the accuracy of genetic diagnosis, and will be most serious in small families which generate limited or no information for or against linkage. In this study 3 out of 35 families are unlinked, and therefore the risk of error for a small family with phase-known flanking markers is approximately 10%. In families large enough to show clear linkage to 16p, the accuracy of diagnosis is 99% using flanking markers (Breuning *et al.*, 1990b). In families that are clearly unlinked to PKD1, the risk for any individual is identical to the statistical risk following a negative ultrasound at any given age. Caution should be exercised if recombination is observed in a family in the absence of flanking marker exchange.

The presence of genetic heterogeneity in ADPKD also has clear implications for molecular studies. A genetically homogeneous population must be identified in order to accurately localise PKD1, and to screen for mutations in any candidate genes. In order to identify the chromosomal site of the PKD2 locus large unlinked families are required, although the three unlinked families identified in this study (PK29, PK52, PK53) will still be of value in confirming linkage and in phenotype-genotype studies.

3.4.2 Two-point linkage analysis

The lod scores between PKD1 and the marker loci are all highly significant, confirming the close linkage of PKD1 to this region. A sex-averaged recombination fraction of 0.07 for PKD1-3'HVR is consistent with previously published results (Reeders *et al.*, 1985; Breuning *et al.*, 1990a). Comparison of the recombination fraction between all pairs of loci should, in theory, allow the order to be determined. In practice, however, for loci separated by less than 10% recombination, too few crossovers are available for precise estimation even when large numbers of meioses are analysed. Multipoint analysis, in which information from multiple loci is considered simultaneously, is therefore used to determine the order of loci (Lathrop *et al.*, 1984).

Two-point analyses were performed between pairs of marker loci. In general, the results are as expected, showing that the markers are closely linked with each other. Four pairs of loci give higher than expected values for θ_{max} [3'HVR (D16S85) and 26-6 (D16S125) ($\theta = 0.16$), 3'HVR (D16S85) and VK5 (D16S94) ($\theta = 0.11$), SM7 (D16S283) and 26-6 (D16S125) ($\theta = 0.17$), and SM7 (D16S283) and VK5 (D16S94) ($\theta = 0.11$)], although the confidence limits ($Z_{max} \pm 1$) are wide and the sample sizes small. Also, the lod score values are low and no conclusions regarding the order of the markers, or the likely physical distance between them, can be made from these data.

3.4.3 Multipoint linkage analysis

Breuning *et al.* (1987b) established flanking markers for PKD1, namely 3'HVR (D16S85) on the distal side and 24-1 (D16S80) proximally. Both markers are separated from PKD1 by a recombination fraction of approximately 0.05. Previous multipoint analyses of the PKD1 region have been described by Reeders *et al.* (1988), Breuning *et al.* (1990a), and Germino *et al.* (1990). From these studies it is clear that PKD1 lies proximal to 3'HVR (D16S85) and 2BP5 (D16S21), and distal to 24-1 (D16S80), CRI-0327 (D16S63), and CRI-090 (D16S45). Further refinement of the map of this region has come from the combined use of linkage analysis and

human-rodent hybrid cell lines. All the probes used by Reeders *et al.* (1988) are distal to the CY13 breakpoint and thus lie in the region 16p13.11-16pter. Within this region, four subregions have been distinguished by means of the 23HA, N-OH1, and CY14 breakpoints. The most important polymorphic loci are thus separated into the following four groups, extending distally: (1) CRI-090 (D16S45), CRI-0327 (D16S63) and 24-1 (D16S80) proximal to 23HA, (2) VK5 (D16S94), 26-6 (D16S125) and SM7 (D16S283) proximal to N-OH1, distal to 23HA, (3) CMM65 (D16S84) and GGG1 (D16S259) proximal to CY14, distal to N-OH1, and (4) EKMDA (D16S83), 2BP5/Fr3-42/cos2B/HMJ1 (all D16S21), and 3'HVR (D16S85) distal to CY14 (Figure 3.1).

Although PKD1 clearly lies proximal to the markers in group 4 and distal to those in group 1, its genetic relationship to the loci within groups 2 and 3 remains less clear. Most of these loci show few, if any, recombinations with PKD1, which on the one hand supports the view that they lie closer to the gene than those in groups 1 and 4, but on the other hand makes ordering difficult. In addition, several pitfalls exist which can result in a misleading genetic assignment, so that it is important to accumulate as much genetic evidence as possible. Firstly, false positive and false negative diagnoses can lead to misinterpretation of results. It is generally accepted that false positive diagnosis by ultrasound scanning is rare, but it has been reported (Reeders *et al.*, 1988). The probability of a false negative diagnosis following an ultrasound scan is age dependent, but very unlikely after the age of 30 years (Parfrey *et al.*, 1990; Coto *et al.*, 1992). Secondly, double recombinants, although very rare in such small genetic distances, can occur and may be mimicked by gene conversion events. Thirdly, mistyping can occur, so that it is important that each apparent recombinant is checked, preferably by an independent laboratory or by typing with other markers. Finally, non-paternity can arise and lead to erroneous interpretation of results, so that this should be checked if possible by genetic fingerprinting.

The results of this study confirm the tight genetic linkage of CMM65 (D16S84) in region 3 which shows only a single recombinant with PKD1 (θ_{max} =

0.03; Z_{max} = 8.37); similarly with VK5 (D16S94) in region 2 which shows no definite recombination with PKD1 (θ_{max} = 0.0; Z_{max} = 8.67). Multipoint analysis shows the most likely order to be D16S84-0.004-PKD1-0.006-D16S283-0.005-D16S125-0.005-D16S94-cen, which gives a multipoint lod score of 21.64.

3.4.4 Recombinants with PKD1

A meiosis showing a single recombination event between the closest informative flanking markers contains information that is useful in ordering markers that lie close to PKD1. In this set of 32 PKD1 families (excluding PK29, PK52 and PK53), 13 single recombinants were identified between PKD1 and the flanking markers. Nine of these recombinants place PKD1 proximal to 3'HVR (D16S85), which is in keeping with the results of the multipoint analysis in this study, and with previous studies (Breuning *et al.*, 1987b and 1990a; Reeders *et al.*, 1988; Germino *et al.*, 1990). Three recombinants place PKD1 distal to 26-6 (D16S125): these are individuals II-1 and II-3 from family PK27, both of whom are clearly affected in a kindred with typical clinical features of ADPKD, and individual III-1 from family PK31 who is unaffected at age 31 years. The presence of recombinants with 26-6 (D16S125) in two families that are not unequivocally PKD1 raises some doubts as to their validity, although in each case exchange of flanking markers has occurred (Figure 3.12; Figure 3.16). Confirmation of these recombinants with another proximal marker has so far not been possible. The final recombination event in individual III-4 of family PK19 places PKD1 proximal to CMM65 (D16S84). This kindred also shows typical features of ADPKD, but III-4 is unaffected with a clearly negative ultrasound scan at age 23 years. At this age, the probability of detecting cysts in a gene carrier by ultrasound scanning is 0.92 (Bear *et al.*, 1992). The possibility therefore exists that he may yet develop ADPKD, but it is more likely that he will remain unaffected. These four recombinants (II-1 and II-3 from PK27; III-1 from PK31; III-4 from PK19) therefore localise the PKD1 gene to the interval between CMM65 (D16S84) distally and 26-6 (D16S125) proximally.

Somlo *et al.* (1992b) identified eleven single recombinant meiotic events in eight linked families, out of over 200 families originally studied. Of these eleven recombinants, three placed PKD1 proximal to GGG1 (D16S259) and two placed PKD1 distal to 26-6 (D16S125). Since GGG1(D16S259) is located approximately 10 kb distal to CMM65 (D16S84) (Germino *et al.*, 1990), the recombinants identified in this study place PKD1 in exactly the same interval as that obtained by Somlo *et al.* (1992b). This interval is known to span approximately 750 kb of DNA (Germino *et al.*, 1992), and within this interval there is some doubt regarding the localisation of PKD1 as defined by recombinants. Somlo *et al.* (1992b) found that the two recombinants locating PKD1 distal to 26-6 (D16S125) also placed the disease distal to the marker 92.6SH1.0, but they also identified a recombinant which located PKD1 proximal to 92.6SH1.0. This individual is unaffected at the age of 33 years, which leaves open the possibility that he may yet develop renal cysts. This seems unlikely given that two studies have shown that all renal cysts in gene carriers are detectable by ultrasound by the age of 30 years (Parfrey *et al.*, 1990; Coto *et al.*, 1992). However, the brother of this individual, who should be affected according to flanking markers, remains unaffected at the age of 40 years, and it may be that this family has a late-onset form of the disease due to allelic variation. Various possibilities exist to explain this apparent anomaly. The family may be unlinked, and this would be consistent with a later than expected onset of the disease. There may have been a gene conversion event, the PKD1 gene may straddle the 92.6SH1.0 locus, there may be more than one gene in the region, or the individual may have a false negative diagnosis. If the gene does straddle the 92.6SH1.0 locus, the three recombinants would be intragenic, suggesting a recombinational hot spot. There are no recombinants in the families analysed in this study which throw any further light on this problem.

3.4.5 New mutations

Dalgaard (1957) calculated the rate of new mutations in ADPKD to be between 6.5 x 10^{-5} and 12 x 10^{-5} per gene per generation, one of the highest mutation rates in

genetic disease. This is likely to be an over-estimate since many patients with ADPKD are asymptomatic with normal renal function, but can be diagnosed on ultrasound scan. Since this technology was not available in 1957, it is likely that many parents have been given a false-negative diagnosis, thereby increasing the apparent mutation rate. As discussed in the following chapter, a high mutation rate is inconsistent with the presence of linkage disequilibrium, which has been detected in this population (see Chapter 4). Only two new mutations have been seen in the 36 families studied, consistent with a low mutation rate and a high reproductive fitness in ADPKD (Milutinovic *et al.*, 1983). Direct determination of the mutation rate from this data is difficult, however, because (i) the size of the population from which the 36 families were ascertained is uncertain because of incomplete ascertainment; (ii) other families with only a single affected member were excluded from the study.

3.4.6 Recombination ratios

In the heterogeneity analysis, LINKMAP was run with probe data from the 3'HVR (D16S85), CMM65 (D16S84), 26-6 (D16S125) and VK5 (D16S94) loci using different fixed ratios of θ_m : θ_f. A broad likelihood peak at a ratio of 5 was obtained, which is consistent with θ_m : θ_f = 6 between 3'HVR (D16S85) and PKD1 (Reeders *et al.*, 1987) and a θ_m : θ_f of 14 between 3'HVR (D16S85) and CRI-090 (D16S63) (Keith *et al.*, 1990). Breuning *et al.* (1990a) also noted a striking excess of θ_m over θ_f across the distal part of their map (3'HVR (D16S85) to CRI-0133 (D16S58). Harris *et al.* (1990) estimated the physical distance between 3'HVR (D16S85) and CMM65 (D16S84) and found that the elevated θ_m :θ_f ratio was due to an increased level of male recombination, which was at least four times greater than that expected in the genome as a whole.

In the present study, 29 single recombinant events were identified in the PKD1 region, with an overall male to female ratio of 2:1. The results, however, suggested a reversal of the ratio from a clear male excess distally to a female excess

towards VK5 (D16S94). In addition, during checking of the MLINK two-point analyses between markers using the ILINK program, a surprising difference in the θ_m :θ_f ratio was obtained for the different intervals. A ratio of 5.0 was obtained for the interval 3'HVR (D16S85)-CMM65 (D16S84), a finding entirely in keeping with the above published data. In the intervals CMM65 (D16S84)-SM7 (D16S283) and SM7 (D16S283)-26-6 (D16S125) however, there was a marked reversal of the ratio with $\theta_f : \theta_m$ being 20.0 and 30.0 respectively. The ratio then fell again to $\theta_f : \theta_m$ = 0.2 in the final interval between 26-6 (D16S125) and VK5 (D16S94). Similar results were obtained with three-point analyses using LINKMAP, although in this case the excess of female over male recombination extends to beyond VK5 (D16S94), where it was again found to be falling. The corresponding likelihood lod scores are highly significant (ranging from 7.31 to 32.02 in the LINKMAP calculations, and 2.84 to 10.97 in the ILINK calculations: see Tables 3.7 and 3.8).

In this study, the number of meioses during which recombination has occurred is very small, and the results therefore have to be treated with caution. Nevertheless there is a consistent trend towards reversal of the ratio from a male excess in recombination distally to a female excess proximal to CMM65 (D16S84). This trend is seen both in the linkage programs, ILINK and LINKMAP, and in analysis of the recombinants themselves. 16p13.3 is located within an R band on Giemsa staining, and falls into the type of R band known as a T band (Holmquist, 1992). T bands contain the highest number of CpG islands, are very GC-rich and have a high gene density (Holmquist, 1992; Craig and Bickmore, 1993). Meiotic chiasmata are the cytological manifestation of crossing over, and these occur very frequently in T bands (Holmquist, 1992). Since R band regions contain less densely packed chromatin (Craig and Bickmore, 1993) they may be more open for recombination to occur.

The overall length of the female map of chromosome 16 exceeds that of the male by 78 cM (Keith *et al.*, 1990), consistent with the fact that the genetic maps of female autosomes are approximately 90% longer than those of the male (Donis-Keller *et al.*, 1987). There are, however, certain chromosomal regions within which

males exhibit higher recombination rates than females, and these appear to be preferentially located in distal chromosomal regions (Keith *et al.*, 1990). All the published data summarised above have analysed regions extending considerably outside the PKD1 region itself, and there is little data available for the recombination ratios in each interval. Indeed, Germino *et al.* (1990) comment that, in their study of 201 meioses, too few recombination events were seen either to make an estimation of $\theta_m : \theta_f$ for smaller intervals meaningful or to allow comparison of $\theta_m : \theta_f$ for adjacent intervals within the array. Although this may also be true for this study, the observation of a change in the $\theta_m : \theta_f$ ratio may be of interest, and may be associated, for example, with the presence of differential male or female gamete-specific expression of genes within the region.

Conclusions

In summary, the results presented in this chapter confirm the presence of genetic heterogeneity in ADPKD, and estimate the proportion of linked families (PKD1) to be 81% in this Scottish population. Two-point analyses between PKD1 and marker loci confirm the close linkage of PKD1 to the region of chromosome 16p flanked by the 3'HVR (D16S85) and VK5 (D16S94) loci, and multipoint analysis shows the most likely order to be D16S84-0.004-PKD1-0.006-D16S283-0.005-D16S125-0.005-D16S94-cen. Four recombinants have been identified which are consistent with a localisation of PKD1 to the interval CMM65 (D16S84)-26-6 (D16S125). Marked variability of the female : male recombination ratios in the different intervals between 3'HVR (D16S85) and VK5 (D16S94) has been noted, but the significance of this is unclear.

CHAPTER FOUR

GENETIC MAPPING II

4.1 LINKAGE DISEQUILIBRIUM

If the alleles at two loci are associated at random in a population (linkage equilibrium), then the frequency of the combination of any two alleles will equal the product of the individual allele frequencies. Thus, at equilibrium, if the frequency of allele A at one locus is 0.5 and the frequency of allele B at a second locus is 0.3, then the frequency of the gametes carrying both A and B would be 0.15 (Lewontin, 1988).

Linkage disequilibrium, or non-random association, refers to the situation where alleles at two loci may be found together more or less often than would be predicted from their frequency in the general population. Genetic recombination is a major factor acting to randomise the alleles at different linked loci, and thus linkage disequilibrium may reflect the rarity of crossing over in meiosis between the two loci, and therefore the long period of time required to reach genetic equilibrium for very closely linked loci. Factors acting to produce non-random association of loci include natural selection, migration, mixing of genetically dissimilar populations (population stratification), genetic drift, small population size and non-random mating (such as inbreeding and assortative mating).

For linkage disequilibrium to be detected between two loci, such as a genetic marker and a disease mutation, the markers are usually closely linked and have low mutation rates. If two loci are separated by more than 10^5 base pairs, the effects of recombination in producing random association render the effects of mutation negligible, and for loci located more closely together the effects of mutation and recombination are likely to be comparable and of less importance than other factors such as those mentioned above (Carothers and Wright, 1992). Assuming that mutation rates do not exceed 10^{-4} per gamete per generation, these considerations would exclude mutation as a general explanation for anomalous

situations where linkage disequilibrium exists between two linked loci, such as *Z* and *A*, but not between *Z* and *B*, where *B* is a third locus known to be closely linked to *A* (Carothers and Wright, 1992). Huntington's Disease is such a situation (MacDonald *et al.*, 1991).

The presence of linkage disequilibrium, as a genetic mapping tool, has been of value in the search for two recently cloned genes, namely those responsible for cystic fibrosis and Huntington's disease (Estivill *et al.*, 1987; Snell *et al.*, 1989). In cystic fibrosis, strong linkage disequilibrium was noted for closely linked DNA markers with little or none detected for the more distant markers, which thus provided genetic confirmation of the location of the gene. These studies also provided information on the origin and homogeneity of the cystic fibrosis (CF) defect, with the unusually high degree of linkage disequilibrium suggesting that a substantial proportion of CF mutant genes in the population are descended from a single mutation (Estivill *et al.*, 1987). In Huntington's disease, a complex pattern of linkage disequilibrium was obtained, perhaps suggesting that multiple mutations had occurred (MacDonald *et al.*, 1991). Nevertheless, the finding of strong linkage disequilibrium between the disease and DNA markers was able to support a localisation of the gene close to these markers. The alternative method of genetic fine mapping is by haplotype analysis of individuals recombinant for the disease (see Chapter 3). Although recombinational events can be useful in providing evidence in favour of a certain locus order they can be misinterpreted due to misdiagnosis, gene conversion or the assumption of a single recombinant as the simplest explanation of a haplotype. The advantage of disequilibrium mapping over the single crossover approach, as one gets closer to the gene, is that it enables data from all families to be used, rather than relying on a decreasing number of informative recombinational events. The potential disadvantage is that a "plateau" of linkage disequilibrium may be reached covering several hundred kilobases of DNA, within which further refinement of the localisation is impossible (Kerem *et al.*, 1989).

4.2 ASSESSMENT OF LINKAGE DISEQUILIBRIUM

As described in Materials and Methods, ADPKD families were ascertained and their members typed for the markers 3'HVR (D16S85), CMM65 (D16S84), SM7 (D16S283) 26-6 (D16S125), and VK5 (D16S94). Three families found to be unlinked to chromosome 16 in the heterogeneity analysis were excluded (PK29, PK52, PK53).

The results were transferred to the pedigrees and haplotypes constructed. In twenty-nine families the haplotypes could be determined unambiguously, and in the remaining four, the haplotypes were determined by minimising the number of crossovers. The affected chromosome in each family was established in this way and allele frequencies or marker to marker haplotype frequencies calculated. The normal chromosomes used to establish control frequencies were derived from the unaffected spouses of family members. Because of the distance of alpha-globin from PKD1, the 3'HVR results were used to help establish haplotypes but were not analysed for the presence or absence of linkage disequilibrium.

4.3 RESULTS

4.3.1 Haplotypes

The haplotypes of affected (PKD1) and normal chromosomes for each family are displayed in Tables 4.1 and 4.2.

4.3.2 PKD1 - marker linkage disequilibrium studies

The presence or absence of allelic association was determined by establishing the frequency of marker alleles in haplotypes of PKD1 chromosomes compared with the control population and applying a chi-squared test. In the case of the multi-allelic microsatellite marker, SM7 (D16S283), all cells with values less than 5 were pooled prior to analysis. χ^2 values were derived using Yates' correction. The results of the analysis are shown in Tables 4.3 to 4.6.

Table 4.1 HAPLOTYPES OF PKD1 CHROMOSOMES

Family	CMM65	SM7	26-6	VK5
PK 1	2	+6	2	2
PK 2	1	0	1	2
PK 3	1	+2	2	2
PK 4	2	0	2	2
PK 5	2	0	1	2
PK 6	1	0	2	2
PK 7	2	-2	2	2
PK 8	1	0	2	2
PK 11	2	+6	2	2
PK 17	1	+4	2	1
PK 18	1	0	1	2
PK 19	2	0	2	2
PK 24	1	0	-	2
PK 25	2	-2	2	1
PK 26	1	0	2	1
PK 27	-	0	2	-
PK 28	2	-6	2	2
PK 30	2	+4	-	2
PK 31	2	+6	2	2
PK 32	-	-2	1	2
PK 35	1	0	1	1
PK 36	2	0	1	2
PK 37	-	-12	1	-
PK 38	1	0	-	2

Table 4.1 Continued

Family	CMM65	SM7	26-6	VK5
PK 39	1	0	-	2
PK 41	1	-2	1	2
PK 42	-	+4	1	2
PK 45	2	0	2	1
PK 49	2	0	2	1
PK 50	1	-	-	2
PK 55	2	+6	-	-
PK 56	1	-2	2	-

Table 4.2 HAPLOTYPES OF NORMAL CHROMOSOMES

Family	CMM65	SM7	26-6	VK5
PK 1	2	+8	1	2
	2	0	2	1
PK 2	2	+2	2	1
	2	+10	2	1
	1	-	2	2
	2	-	1	2
	2	-	2	1
	2	-	2	1
PK 3	1	-4	2	2
	1	0	1	2
	1	-	2	1
	2	-	1	1
PK 4	1	0	2	2
	1	+2	1	1
	1	-	2	2
	1	-	1	1
PK 5	1	0	2	2
	1	0	2	1
PK 6	2	-4	1	2
	1	0	2	1
PK 7	2	0	1	2
	2	0	2	2
	1	0	2	1
	1	0	2	1
	2	0	2	1

Table 4.2 Continued

Family	CMM65	SM7	26-6	VK5
PK 7	1	0	2	2
	1	-	2	2
	2	-	2	2
PK 8	1	0	2	1
	2	0	2	2
PK 11	2	-	2	2
	2	-	2	2
	1	+6	2	2
	1	0	2	2
	2	0	1	2
	1	+2	2	1
PK 17	1	0	2	1
	2	0	1	2
PK 18	2	0	2	1
	2	0	2	2
	1	+6	2	-
	2	+2	2	-
PK 19	2	0	1	1
	1	0	2	1
PK 24	1	0	-	2
	1	+2	-	2
PK 25	2	0	1	1
	2	0	2	1
PK 26	1	0	2	1
	1	0	2	1

Table 4.2 Continued

Family	CMM65	SM7	26-6	VK5
PK 27	-	0	-	1
	-	-8	-	2
PK 28	2	0	1	1
	2	0	2	2
PK 29	-	0	1	2
	-	+2	2	2
	-	0	1	1
	-	0	1	2
PK 30	1	-2	-	2
	2	0	-	1
	2	-	-	1
	2	-	-	1
PK 31	2	-6	2	1
	2	0	1	2
PK 32	-	-2	2	1
	-	+4	2	1
PK 35	1	0	2	1
	1	+8	2	2
PK 36	2	+6	2	1
	2	+4	2	1
PK 37	-	0	2	-
	-	0	2	-
	-	-	1	2
	-	-	2	2

Table 4.2 Continued

Family	CMM65	SM7	26-6	VK5
PK 38	1	0	-	1
	2	+2	-	2
PK 41	1	-	2	1
	2	-	2	1
PK 42	-	0	2	2
	-	-6	2	2
PK 45	2	-6	-	-
	2	+6	-	-
PK 49	2	0	2	-
	2	0	2	-
	2	-2	2	1
	2	0	2	2
PK 50	1	-	-	2
	2	-	-	2

Table 4.3 CMM65 (D16S84)

	Affected	Control	Total	χ^2	p
Allele 1	14	32	46		
Allele 2	14	42	56		
	28	74	102	0.151	n.s.

Table 4.4a SM7 (D16S283)

	Affected	Control	Total
Allele 1 (+10)	0	1	1
Allele 2 (+8)	0	2	2
Allele 3 (+6)	4	4	8
Allele 4 (+4)	3	2	5
Allele 5 (+2)	2	7	9
Allele 6 (0)	15	43	58
Allele 7 (-2)	5	3	8
Allele 8 (-4)	0	2	2
Allele 9 (-6)	1	3	4
Allele 10 (-8)	0	1	1
Allele 11 (-12)	1	0	1
	31	68	99

Table 4.4b SM7 (Contracted)

	Affected	Control	Total	χ^2	p
Alleles 1-5	9	16	25		
Allele 6	15	43	58		
Alleles 7-11	7	9	16		
	31	68	99	1.116	n.s.

Table 4.5 26-6 (D16S125)

	Affected	Control	Total	χ^2	p
Allele 1	9	18	27		
Allele 2	17	56	73		
	26	74	100	0.578	n.s.

Table 4.6 VK5 (D16S94)

	Affected	Control	Total	χ^2	p
Allele 1	6	40	46		
Allele 2	22	40	62		
	28	80	108	5.814	**< 0.02**

No evidence of linkage disequilibrium was found between PKD1 and markers CMM65 (D16S84), 26-6 (D16S125) and SM7 (D16S283), all of which show close genetic linkage to PKD1 (Breuning *et al.*, 1990a; Harris *et al.*, 1991). In the case of VK5 (D16S94), also closely linked (Hyland *et al.*, 1990), there is evidence of linkage disequilibrium: the affected haplotypes in 6 kindreds showing the 1.6 kb allele (allele 1) and in 22 kindreds the 1.3 kb allele (allele 2), compared with the control

population who showed 40 1.6 kb alleles and 40 1.3 kb alleles. This result is significant at the 2% level (Yates' corrected χ^2 = 5.814). The allele frequencies in the control population are in good agreement with those published previously (Breuning *et al.*, 1990a; Harris *et al.*, 1991).

4.3.3 Marker - marker disequilibrium studies

In order to assess whether or not there was any disequilibrium between DNA markers, the allele frequencies for markers on the normal chromosomes were first established. Pairs of markers were then compared using Fisher's exact test (two-tailed). For pairs of markers where all the cells contained five or more, the chi-squared test using Yates' correction was also performed, and the results from the two tests compared. The results are shown in Tables 4.7 to 4.12.

Table 4.7 CMM65 (D16S84) - SM7 (D16S283)

CMM65 / SM7	Allele 1	Allele 2	Total	χ^2	p	Fisher's exact test (two-tailed)
Allele 0	17	19	36			
Other alleles	8	12	20			
Total	25	31	56	0.05	n.s.	p = 0.78

Table 4.8 CMM65 (D16S84) - 26-6 (D16S125)

CMM65 / 26-6	Allele 1	Allele 2	Total	χ^2	p	Fisher's exact test (two-tailed)
Allele 1	3	11	14			
Allele 2	24	24	48			
Total	27	35	62	-	-	p = 0.07

Table 4.9 CMM65 (D16S84) - VK5 (D16S94)

CMM65 / VK5	Allele 1	Allele 2	Total	χ^2	p.	Fisher's exact test (two-tailed)
Allele 1	16	20	36			
Allele 2	15	17	32			
Total	31	37	68	0.002	n.s.	p = 1.00

Table 4.10 SM7 (D16S283) - 26-6 (D16S125)

26-6 / SM7	Allele 1	Allele 2	Total	χ^2	p	Fisher's exact test (two-tailed)
Allele 0	11	28	39			
Other alleles	3	16	19			
Total	14	44	58		n.s.	p = 0.35

Table 4.11 SM7 (D16S283) - VK5 (D16S94)

VK5 / SM7	Allele 1	Allele 2	Total	χ^2	p	Fisher's exact test (two-tailed)
Allele 0	21	18	39			
Other alleles	10	11	21			
Total	31	29	60	0.036	n.s.	p = 0.79

Table 4.12 26-6 (D16S125) - VK5 (D16S94)

VK5 26-6	Allele 1	Allele 2	Total	χ^2	p	Fisher's exact test (two-tailed)
Allele 1	7	11	18			
Allele 2	28	22	50			
Total	35	33	68	0.937	n.s.	p = 0.28

No evidence of statistically significant linkage disequilibrium was detected between markers using this method of analysis. Haplotypes obtained for the affected chromosomes were then compared with those for the control chromosomes, using the chi-squared test. Once more, cells containing numbers less than 5 were pooled, and values were obtained using Yates' correction. The results are shown in Tables 4.13 to 4.18.

Table 4.13 CMM65 (D16S84) - SM7 (D16S283)

	Affected	Control	Total	χ^2	p
Haplotype 1/0	8	17	25		
Haplotype 2/0	6	19	25		
All other haplotypes	12	20	32		
	26	56	82	2.603	n.s.

Table 4.14 CMM65 (D16S84) - 26-6 (D16S125)

	Affected	Control	Total	χ^2	p
Haplotype 1/1 + 2/1	5	13	18		
Haplotype 1/2	5	25	30		
Haplotype 2/2	10	24	34		
	20	62	82	0.838	n.s.

Table 4.15 CMM65 (D16S84) - VK5 (D16S94)

	Affected	Control	Total	χ^2	p
Haplotype 1/1 + 2/1	6	35	41		
Haplotype 1/2	11	15	26		
Haplotype 2/2	9	16	25		
	26	66	92	5.26	< 0.1

Table 4.16 SM7 (D16S283) - 26-6 (D16S125)

	Affected	Control	Total	χ^2	p
Haplotype 0/2	9	28	37		
All other haplotypes	18	30	48		
	27	58	85	1.118	n.s.

Table 4.17 SM7 (D16S283) - VK5 (D16S94)

	Affected	Control	Total	χ^2	p
Haplotype 0/2	12	18	30		
All other haplotypes	16	43	59		
	27	61	89	0.988	n.s.

Table 4.18 26-6 (D16S125) - VK5 (D16S94)

	Affected	Control	Total	χ^2	p
Haplotype 1/1 + 1/2	7	18	25		
Haplotype 2/1	5	27	32		
Haplotype 2/2	11	21	32		
	23	66	89	2.06	n.s.

4.3.4 Assessment of linkage disequilibrium with other markers

In view of the finding of linkage disequilibrium with VK5 (D16S94), and the potential implications of this finding for the genetic localisation of PKD1, it was decided to look for evidence of linkage disequilibrium with a series of other markers. Aliquots of DNA from each family were sent to Dr. Peter Harris (Oxford) who typed them for four new microsatellite markers in the region, namely SM6 (2), CW4 (D16S291), CW3, and CW2 (Figure 3.3). The results of the analysis are shown in Tables 4.13 and 4.14. There is evidence of linkage disequilibrium with marker CW4 (D16S291), but not with any of the other three markers.

Aliquots of DNA were also sent to Angela Snarey (London) who typed them for ten markers in the region, five of which had not been previously used to type these families. Both cKLH9, which is the same as CW4 (D16S291), and a new marker, W5.2 (CW1), showed statistically significant linkage disequilibrium with PKD1: the other eight markers were not in linkage disequilibrium with PKD1. Both cKLH9/CW4 (D16S291) and W5.2 (CW1) are located at least 300 kb proximal to CMM65 (D16S84) (see Figure 3.3).

Attempts to construct extended haplotypes using data from all the typed markers in this set of families has failed to identify a single founder chromosome (A. Snarey, personal communication; manuscript submitted). This suggests the presence of more than one independent mutations within the Scottish PKD1 families.

Table 4.19 ALLELE FREQUENCIES IN EDINBURGH PKD1 FAMILIES

SM6 (2)						CW2				
	Normal		PKD1				Normal		PKD1	
Allele	No.	%	No.	%		Allele	No.	%	No.	%
+ 16	1	.8	0	0		+ 16	1	0.8	0	0
+ 14	3	2.3	2	7.7		+ 10	1	0.8	0	0
+ 12	3	2.3	0	0		+ 8	0	0	1	4.8
						+ 6	5	4.0	0	0
+ 4	2	1.6	1	3.8		+ 4	12	9.6	4	15.4
+ 2	28	22.7	6	23.0		+ 2	46	36.8	11	42.3
0	54	41.9	9	34.6		0	32	25.6	4	15.4
- 2	1	0.8	0	0		- 2	14	11.2	0	0
- 4	1	0.8	1	3.8		- 4	1	0.8	0	0
- 6	16	12.4	4	15.4		- 6	1	0.8	0	0
- 8	7	5.4	2	7.7		- 8	1	0.8	0	0
- 10	8	6.2	0	0		- 12	2	2.0	0	0
- 14	1	0.8	0	0		- 14	10	8.0	6	23.1
- 30	1	0.8	0	0						
- 32	3	2.3	1	3.8						

Allele frequencies for PKD1 and normal chromosomes in the Scottish population. Families typed for markers SM6 (2) and CW2 by P. Harris (Oxford).

Table 4.20 ALLELE FREQUENCIES IN EDINBURGH PKD1 FAMILIES

CW4						CW3				
	Normal		PKD1				Normal		PKD1	
Allele	No.	%	No.	%		Allele	No.	%	No.	%
+ 6	2	2.2	1	4.8		+ 10	0	0.8	0	0
+ 4	2	2.2	5	***23.8**		+ 6	3	2.4	1	3.8
+ 2	35	38.5	3	***14.3**		+ 4	31	25.2	6	23.0
0	16	17.6	2	9.5		+ 2	19	15.4	4	15.3
- 2	15	16.5	3	14.3		0	68	55.3	12	46.2
- 4	4	4.4	1	4.8		- 2	0	0	2	7.7
- 6	17	18.7	5	23.8		- 4	1	0.8	0	0

*** p = <0.05**

Allele frequencies for PKD1 and normal chromosomes in the Scottish population. Families typed for markers CW4 and CW3 by P. Harris (Oxford).

4.4 DISCUSSION

These results suggest the presence of linkage disequilibrium between PKD1 and VK5 (D16S94) in a Central Scottish collection of 32 ADPKD families. Any initial disequilibrium occurring when a disease mutation arises is dissipated at the rate of $(1-r)^n$ in n generations, where r is the recombination fraction separating marker and disease locus. If r is very small, the rate of approach to equilibrium is slow. Similarly, a recent mutation will take a finite time to reach equilibrium with adjacent markers, even when r is significant. Since the recombination fraction between PKD1 and VK5 is expected to be very small, (θ_{max} = 0.00 on two-point linkage analysis in this population - see Chapter 3), any mutation arising is likely to be in disequilibrium for some considerable time. As discussed in the previous chapter, the identification of meiotic crossover points by haplotype analysis has led to the proposal that PKD1 lies between CMM65 (D16S84) and 26-6 (D16S125), a region shown by pulsed field gel electrophoresis to contain less than 750 kb of DNA and predicted to contain a large number of genes (Germino *et al.*, 1992). The more proximally located marker VK5 (D16S94) lies just outside this area. The suggestion of linkage disequilibrium between VK5 (D16S94) and PKD1 has implications for the genetic localisation of PKD1 and, if confirmed by other groups with ethnically homogeneous patient populations, it would argue for a detailed physical mapping effort in the proximal part of the CMM65 (D16S84) - 26-6 (D16S125) interval and extending towards VK5 (D16S94). There has been one report describing failure to find linkage disequilibrium between VK5 (D16S94) and PKD1 (Elles, 1992) in 38 families from the North West of England. In this population, the allele frequencies observed for 26-6 (D16S125) and VK5 (D16S94) were significantly different from those seen in the Scottish and Dutch populations (this study; Breuning *et al.*, 1990a). Linkage disequilibrium may be obscured in other populations by allelic heterogeneity at PKD1: either the mutation may have arisen in another population associated with a common marker haplotype or different mutations in the PKD1 gene may be present in the same population. Linkage disequilibrium may be present but remain undetected due to small sample size (Thompson *et al.*, 1988). It

is possible that the finding of linkage disequilibrium in the Scottish population is a chance occurrence associated with a small sample of affected chromosomes, although this is made less likely by the finding of linkage disequilibrium between PKD1 and other markers in the region (see below).

Since the two-point linkage analysis between PKD1 and 26-6 (D16S125) also gave a θ_{max} of 0.00, the failure to find linkage disequilibrium between PKD1 and 26-6 (D16S125) is perhaps surprising. This could be because a common PKD1 mutation arose in association with the more frequent 26-6 allele (allele 2), making it more difficult to detect disequilibrium, or because the polymorphism is itself subject to significant mutational change. Microsatellite markers, such as SM7 (D16S283), are thought to have relatively high mutation rates compared with RFLP markers. Weissenbach *et al.* (1992) estimated that the mutation rate of microsatellites was close to 1×10^{-3} per locus per gamete, and more recent data suggest that it may be as high as 1×10^{-2} per locus per gamete (J. Weissenbach, personal communication). As discussed above, a high mutation rate would serve to mask the presence of any linkage disequilibrium between loci. Despite this possibility, linkage disequilibrium has been demonstrated between two highly polymorphic microsatellite markers on chromosome 5 separated by 7 kb (Sherrington *et al.*, 1991), suggesting that microsatellites may be sufficiently stable to allow this finding to be detected. Again, the most likely explanation for the absence of linkage disequilibrium is that the PKD1 mutation is associated with the most frequent SM7 (D16S283) allele.

In order to investigate further the finding of linkage disequilibrium between PKD1 and VK5 (D16S94), aliquots of DNA from the Scottish families were sent to Dr. Peter Harris in Oxford where they were typed for four markers (SM6 (2), CW4 (D16S291), CW3, and CW2). They found evidence of linkage disequilibrium between PKD1 and CW4 (D16S291) but not with any of the other markers. CW4 lies approximately half-way between CMM65 (D16S84) and 26-6 (D16S125), as can be seen in Figure 3.3. The same aliquots were typed by Ms. A. Snarey for probes KG8, KM17, Blu-2, Nik2.9 and W5.2. Linkage disequilibrium was demonstrated between

PKD1 and W5.2 in the Scottish population, and also confirmed between cKLH9/CW4 (D16S291) and PKD1. Thus, evidence of linkage disequilibrium has been found between PKD1 and each of the markers CW4/cKLH9 (D16S291), W5.2 and VK5 (D16S94), but not with any of the other markers in the same region [CMM65 (D16S84), KG8, SM6 (2), KM17, Blu-2, Nik2.9, CW3, CW2, SM7 (D16S283) and 26-6 (D16S125)]. The fact that linkage disequilibrium has not been detected between PKD1 and all the markers tested may be due to sample size, to the varying abilities of individual loci to detect linkage disequilibrium as a result of allele frequencies, to mutation at individual loci or to the presence of multiple mutations.

In Huntington's disease (HD), linkage disequilibrium was initially detected between HD and both the D4S95 and D4S98 loci (Snell *et al.*, 1989; Theilmann *et al.*, 1989). More extensive studies showed that the pattern of linkage disequilibrium was extremely complicated (MacDonald *et al.*, 1991), and a convincing peak of linkage disequilibrium was never detected. The use of multi-allele systems to construct haplotypes of HD chromosomes showed a wide diversity of haplotypes, but also suggested that a common segment was shared between D4S182 and D4S180 on these chromosomes, a distance of approximately 500 kb (MacDonald *et al.*, 1992). The identification of a gene containing a trinucleotide repeat that is expanded and unstable on HD chromosomes proved to be consistent with this localisation (The Huntington's Disease Collaborative Research Group, 1993). Alternatively, a plateau of linkage disequilibrium can define an area on either side of a gene, as was seen in cystic fibrosis (Kerem *et al.*, 1989). In this case, however, although the association coefficients appeared to be rather constant over a 300 kb interval, fluctuations were detected with several markers not showing linkage disequilibrium within the plateau (Kerem *et al.*, 1989). As in HD, the location of the gene proved to be consistent with this localisation by linkage disequilibrium.

It is therefore possible that a plateau of linkage disequilibrium between CW4/cKLH9 (D16S291) and VK5 (D16S94) may be present in PKD1, with the picture being obscured by various factors such as allelic heterogeneity at PKD1, the varying ability of each marker locus to detect linkage disequilibrium due to allelic

variation and mutation of the microsatellite marker loci. In Huntington's Disease, recombinant mapping led to conflicting predictions with one recombination event suggesting that the most terminal markers flank the disease gene, and two others favouring a telomeric location for the defect (MacDonald *et al.*, 1989). Linkage disequilibrium data was proved to be accurate by the isolation of the gene. Errors in localisation by recombination mapping can occur because there are only a small number of events and several opportunities for error such as mistyping, misinterpretation (e.g. gene conversion, double crossovers) and misdiagnosis. The finding of linkage disequilibrium in PKD1 may therefore be important in the localisation of the gene.

The markers analysed in this study do not show any marker-marker disequilibrium. This is, perhaps, surprising given that SM7 (D16S283), 26-6 (D16S125) and VK5 (D16S94) are separated from each other by only approximately 100 kb (Breuning *et al.*, 1990a; Harris *et al.*, 1991). However, if the marker alleles have low mutation rates and arose a long time ago, there will have been sufficient time for them to reach linkage equilibrium. CMM65 (D16S84) is located 750 kb distal to 26-6 (D16S125) (Germino *et al.*, 1992), so failure to find linkage disequilibrium is less surprising. Previous studies also demonstrate that there is no simple relationship between marker-marker distance and the presence or absence of linkage disequilibrium. Litt and Jorde (1986) found no consistent relationship between the amount of linkage disequilibrium and the physical distance between pairs of loci within a highly polymorphic region of chromosome 2q spanning 20 kb. They suggest that since recombination events are very rare between sites that are only several kilobases apart, other factors, such as mutation, admixture, and drift, may disturb the expected relationship between distance and linkage disequilibrium. Hegele *et al.* (1990) similarly found that five physically close pairs of markers at the low-density lipoprotein (LDL) receptor locus were in significant linkage disequilibrium, but that other physically close markers were in linkage equilibrium with each other. These markers were separated from each other by a maximum of 20 kb. At the β-globin locus, two clusters of linkage disequilibrium, separated by

9.1 kb of DNA, were found to be in equilibrium with each other due to a "hot spot" of recombination between the two clusters (Chakravarti *et al.*, 1984). Finally it has been suggested that all physically close markers may be in linkage disequilibrium that is undetectable because of the sample size and the magnitude of allele frequencies (Thompson *et al.*, 1988). The absence of marker-marker disequilibrium in this study, therefore, may be due to a variety of factors.

The population prevalence of ADPKD is high (1 in 1000), and approximately 80% of cases are the result of mutations in the PKD1 gene (Wright *et al.*, 1993). This might be consistent with a high mutation rate for PKD1, although it should be noted that the reproductive fitness of ADPKD patients is high (Dalgaard, 1957; Milutinovic *et al.*, 1983). The higher the reproductive fitness, the lower the mutation rate required to maintain a prevalence of this magnitude. Linkage disequilibrium is not expected to occur if the mutation rate at PKD1 is high since each new mutation is likely to occur on a different chromosomal background, and equilibrium is reached more rapidly. The results described in these ADPKD families would therefore suggest that the high prevalence is maintained by a high reproductive fitness or other factors, such as reproductive compensation or previous selective advantage, in the presence of an intermediate or a low mutation rate.

In conclusion, linkage disequilibrium mapping suggests that the search for PKD1 should be in the region between cKLH9/CW4, W5.2 and VK5 (D16S94). As in Huntington's Disease, there are multiple haplotypes, presumably associated with different PKD1 mutations. Linkage disequilibrium may therefore be seen with some markers (common to the different mutations), but not with others in the same region.

CHAPTER FIVE

PHYSICAL MAPPING STUDIES

5.1 YEAST ARTIFICIAL CHROMOSOMES

The yeast artificial chromosome (YAC) vector pYAC4 (Figure 5.1) is constructed from a pBR322 backbone, and retains much of the plasmid sequence including the bacterial origin of replication (*ori*) and the ampicillin resistance gene (Amp^r). The vector also has all the essential yeast chromosomal elements to enable replication and segregation in a yeast host: a yeast centromere (*CEN4*) and autonomous replication sequence (*ARS*), and two *Tetrahymena* telomere sequences (*TEL*), one for each vector arm. The two vector 'arms' are derived by restriction endonuclease cleavage of the YAC plasmid vector, and large fragments of restricted DNA are ligated between the arms to form a linear YAC. In addition, the vector arms carry selectable markers (*TRP1* on the left arm and *URA3* on the right arm) to enable positive selection. pYAC4 is one of the most commonly used cloning vectors and has a single cloning site in the middle of the ochre suppresser (*SUP4*) gene, whose interruption by inserted DNA results in red recombinant clones in the presence of the *ade2-1* mutation in the yeast host (Burke *et al.*, 1987). AB1380 is a strain of *Saccharomyces cerevisiae* containing a non-Mendelian genetic determinant ψ^+ that enhances the expression of ochre suppresser tRNAs and causes extremely slow growth in SUP^+ strains. It is commonly used as a transformation host for YACs.

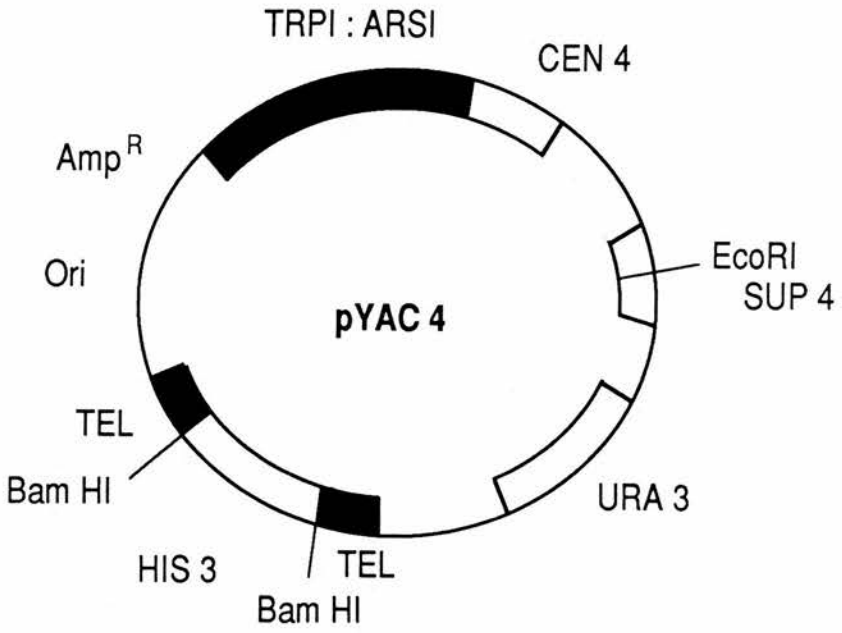
5.1.1 YAC libraries

Three total genomic YAC libraries, two human and one mouse, were used during this project and are discussed below.

ICI (Imperial Chemical Industries) YAC library (Anand *et al.*, 1990)

A YAC library of 35,000 clones was constructed from human lymphoblastoid (48,XXXX) cell line DNA. The restricted DNA fragments were size-fractionated by

Figure 5.1 pYAC4



Yeast artificial chromosome cloning vector pYAC4. The pBR322 sequences are shown as a thin line, and include the ampicillin-resistance gene (*Amp*^{r}) and bacterial origin of replication (*Ori*). *SUP4*, *TRP1*, *HIS3* and *URA3* are yeast genes: *SUP4* is an ochre-supressing allele of a tyrosine tRNA gene that is interrupted when exogenous DNA is cloned into the vector; *TRP1* and *URA3* are present in the artificial chromosomes and allow selection for molecules that have acquired both chromosome arms from the vector; *HIS3* is discarded during the cloning process. *ARS1* and *CEN4* are sequences that are naturally adjacent to *TRP1* on yeast chromosome IV; *ARS1* is an autonomous-replication sequence and *CEN4* provides centromere function. The *TEL* sequences are derived from the termini of the *Tetrahymena* macronuclear ribosomal DNA molecules.

PFGE prior to ligation to the pYAC4 vector, and transformation was performed using the *S. cerevisiae* host strain AB1380. Characterisation of the library shows the average insert size to be 350 kb, giving a 3.5 genome equivalent library. The authors used 14 single copy probes to screen the library by colony hybridisation as well as the polymerase chain reaction (PCR), and isolated between one and five YAC clones per probe. They therefore concluded that there was good overall representation of the human genome in the library.

CEPH (Centre d'Etude du Polymorphisme Humain) YAC library (Albertsen *et al.*, 1990)

This YAC library was constructed from large fragments of DNA from a human lymphoblastoid (46,XY) cell line. Once again, DNA was size-fractionated prior to ligation to the pYAC4 vector, and *S. cerevisiae* AB1380 was used as the yeast host. The library has 50,000 clones with a mean insert size of 430 kb, and contains the equivalent of seven haploid human genomes.

St. Mary's Hospital Medical School mouse YAC library (Chartier *et al.*, 1992)

A mouse YAC library was constructed from female C57BL/10 mice in a recombination-deficient strain of *S. cerevisiae* (3a) carrying a mutation in the RAD52 gene. Size-fractionated DNA was ligated to the pYAC4 vector, and following transformation of the yeast host a library of 41,568 clones was obtained with an average insert size of 240 kb. This represents more than a threefold coverage of the mouse genome.

5.1.2 Screening of YAC libraries

In order to utilise YAC libraries as a source of cloned DNA, large numbers of clones require to be screened in order to detect those containing specific DNA sequences. Colony hybridisation is effective as a screening technique (Brownstein *et al.*, 1989), but it is cumbersome and time-consuming. PCR can be used to assay DNA from pools of YAC clones for the presence of specific sequences (Green and Olson, 1990):

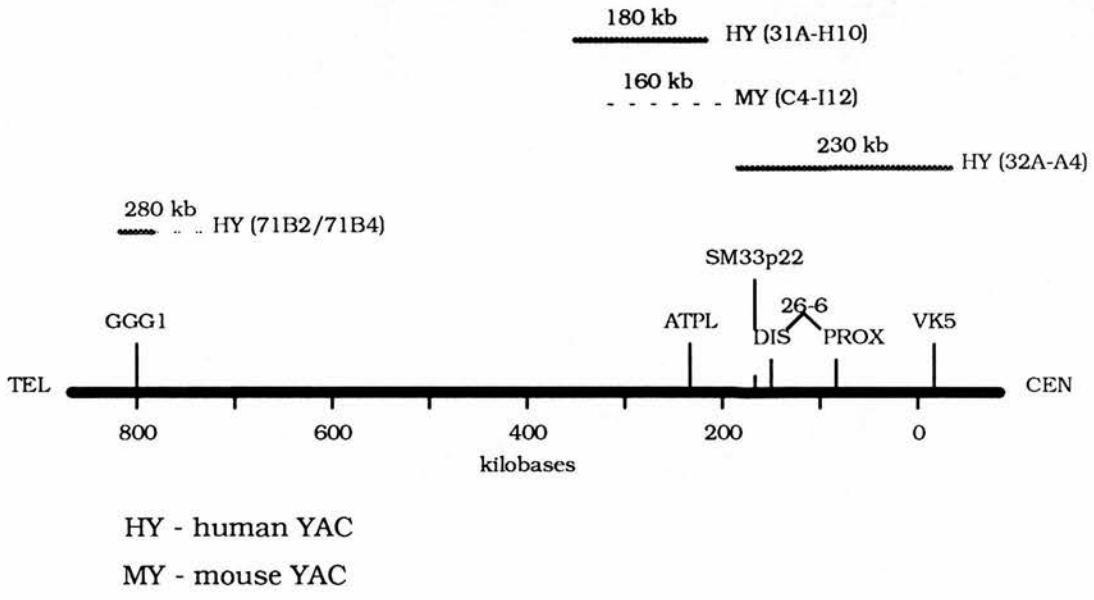
the number of clones in each pool decreases with each round of screening. Several YAC clones may be identified using a single probe. These YACs will have overlapping inserts, all containing the probed locus, and initial alignment of the clones can be done using inter-*Alu* PCR fingerprinting (Nelson *et al.*, 1989). In order to build up a YAC contig containing cloned DNA from a large region, libraries can either be screened with many probes from the area in the hope that the resultant clones will overlap, or YAC walking can be employed whereby sequence information or DNA probes from the end of a YAC are used to re-screen the library for an overlapping clone.

5.2 PHYSICAL MAPPING AROUND PKD1

PCR primers complementary to DNA sequences flanking the PKD1 gene were used to screen YAC libraries in order to obtain cloned DNA from the region. In view of the finding of linkage disequilibrium between PKD1 and VK5 (D16S94), most of the primers used to screen the YAC libraries were in the proximal half of the CMM65 (D16S84)-VK5 (D16S94) interval. PCR primers to the GGG1 (D16S259), SM33p22 and 26-6 (D16S125) loci were made available by Dr. M. H. Breuning (Leiden). In order to improve proximal coverage, it was decided to design primers to the 3' untranslated region of the sequence for AJ1 cDNA, an ATPase-like vacuolar proton channel gene (ATPL) which is located approximately 50 kb distal to 26-6DIS (D16S125) (Gillespie *et al.*, 1991), and to subclone and sequence part of VK5 (D16S94), so that primers could be designed to this locus. The positions of these loci in relation to PKD1 are shown in Figure 5.2.

The aim was to obtain as many YACs as possible from this region, to characterise these YACs, and then use them to screen cDNA libraries in order to identify cDNAs with sequence homology to the cloned DNA. Biotinylated cosmid DNA has been used to enrich PCR-amplified cDNA libraries by hybridisation followed by selection of hybrids using a biotin-streptavidin reaction (Korn *et al.*, 1992). A similar technique using YACs as the source of DNA was thought to be feasible, and this was attempted. The YACs were separated from the host yeast

Figure 5.2



Location of probes GGG1 (D16S259), ATPL, SM33p22, 26-6 (D16S125) and VK5 (D16S94), which were used in screening the YAC libraries. The four YACs (three human and one mouse) obtained are indicated with their approximate positions.

chromosomes by PFGE, digested with a restriction enzyme and the fragments ligated to catch-linkers. This material was then amplified using the 962L primer (complementary to the catch-linker) and re-amplified with the biotinylated 962L primer (C232). cDNA inserts were amplified from the cDNA library by PCR using the λgt10 forward and reverse primers. After pre-competing with *Cot*I DNA, the cDNA inserts and biotinylated YAC DNA were allowed to hybridise and the specific hybrids selected by binding to streptavidin-coated magnetic beads. Non-specific hybrids were removed by washing and the specific hybrids eluted from the beads at various temperatures and amplified using λgt10 primers. This selection process could be repeated as necessary until the hybrids were pure enough to be cloned and sequenced.

The mouse YAC was also used to screen a cosmid library by hybridisation; the identified cosmids could then be used to screen the cDNA libraries directly by filter hybridisation.

5.3 RESULTS

5.3.1 Subcloning and sequencing of VK5 (D16S94) using *Msp*I

Probe VK5 (Hyland *et al.*, 1990) is a 7 kb *Eco*RI-*Bam*HI fragment in pBR322. The plasmid was digested with *Eco*RI and *Bam*HI to release the insert, which was then excised from a 0.8% TAE agarose gel and purified by Geneclean™. Digestion of the insert with *Msp*I yielded 11 bands, the largest of which was 1.3 kb in size. Since *Msp*I has a recognition sequence of CCGG, VK5 must be highly CG-rich. Digestion of pBR322 DNA with *Msp*I gives 26 fragments, the largest of which is 622 bp. pBR322 containing the VK5 insert was digested with *Msp*I, and several bands were obtained, the largest of which was 1.3 kb. This band was excised from agarose, purified and subcloned into pBluescript that had been digested with *Cla*I to give compatible cohesive ends. This DNA was transformed into DH5α cells, which were then plated onto L-agar containing ampicillin, X-gal and IPTG. White recombinant clones were picked, DNA minipreps prepared and the subclone digested with *Eco*RI. The 1.3 kb fragment was cloned in both orientations, and two *Eco*RI sites were

present in each clone (Figure 5.3). Since there are no *Eco*RI sites in VK5, the 1.3 kb fragment must be spanning the pBR322 *Eco*RI cloning site, the other *Eco*RI site being present in the cloning vector. The ends of the 1.3 kb *Msp*I subclone were sequenced using M13 universal and reverse primers, and the ends of the 7 kb VK5 insert were sequenced using primers close to the *Eco*RI and *Bam*HI sites in pBR322. The sequence data obtained are detailed in Appendix 1. One end of the *Msp*I subclone was confirmed as being pBR322 sequence and the other end is unique sequence from VK5. The VK5 sequence adjacent to the pBR322 *Bam*HI cloning site, and the unique VK5 sequence obtained from the *Msp*I subclone, were found to be very GC-rich, with GC contents of 62% and 63% respectively.

5.3.2 VK5/*Msp*I PCR

The internal VK5 sequence together with VK5 end sequence from the *Eco*RI cloning site inwards was used to design forward (VF1) and reverse (VR1) PCR primers using the PRIMER (Version 0.5) program package (Whitehead Institute for Biomedical Research). The primer sequences are as follows:-

VF1 5' GGC CGC GTG TTT TGC TTT TG 3' (Tm = 67.9°C)

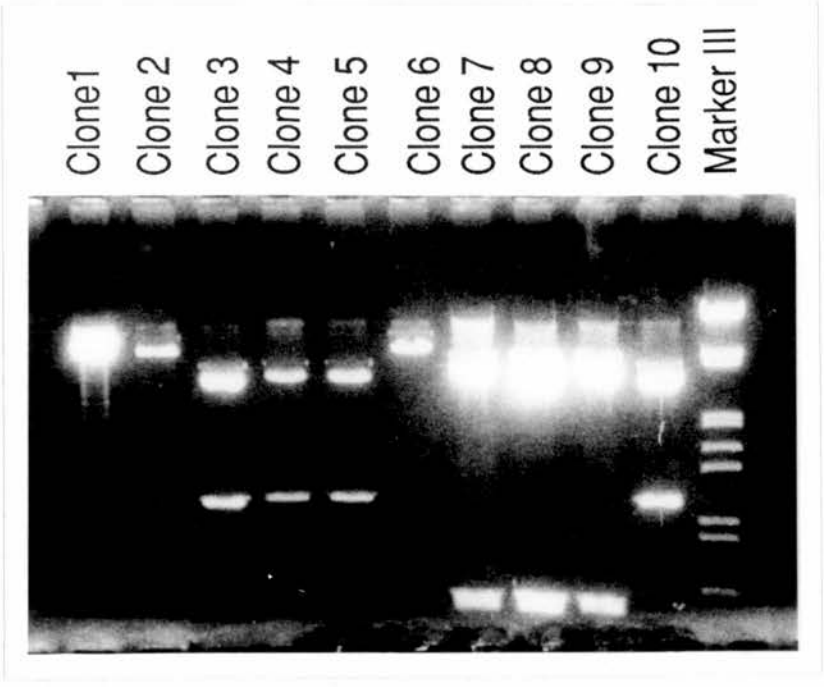
VR1 5' AGA AGG CAA CGA GGG CTG GC 3' (Tm = 68.2°C)

In view of the melting temperatures (Tm) of the primers, an annealing temperature of 63°C was used, and PCR was carried out as described in Materials and Methods. The expected product of approximately 1.3 kb was obtained. The optimal $MgCl_2$ concentration was found to be 1.0 mM; multiple bands were obtained at higher concentrations.

5.3.3 Subcloning and sequencing of VK5 (D16S94) using *Kpn*I

The above subcloning experiment was repeated using *Kpn*I. Digestion of the VK5 plasmid with *Kpn*I gives three bands, approximately 7 kb, 3.5 kb and 1.4 kb in size. Since there are no *Kpn*I sites in pBR322, the 3.5 kb and 1.4 kb fragments must be within the VK5 insert. This was confirmed by digestion of the insert itself with the same enzyme, giving four fragments of approximately 3.5 kb, 2.0 kb, 1.4 kb and

Figure 5.3



*Eco*RI digests of ten randomly-picked VK5/*Msp*I clones. Clones 1, 2, and 6 do not contain inserts. Clones 3, 4, 5 and 10 contain the VK5/*Msp*I insert in the opposite orientation from clones 7, 8 and 9. Two *Eco*RI sites are present in all the clones containing inserts. The cloning vector is pBluescript.

<0.1 kb in size. The 1.4 kb fragment was subcloned into pBluescript KS, which had also been digested with *Kpn*I, by the same method as before, and the ends sequenced (Appendix 1). The ends of this fragment were also found to be very GC-rich, with a GC content of 62-64%. A 28 bp repeat, separated by 25 bp, was seen very close to one end of the fragment.

Primers were designed to the *Kpn*I fragment forward and reverse sequence, avoiding the repeat sequence at the forward end. This was done using the PRIMER (Version 0.5) program. The forward (C719) and reverse (C718) primer sequences are as follows:-

C719 5' CCA GGC CTC AGC ATC CCA AG 3' (Tm = 67.5°C)

C718 5' CAC TGC TGC CCA CTG GGT CA 3' (Tm = 68.2°C)

PCR was initially carried out using the same program as for the VK5/*Msp*I PCR, except that an annealing temperature of 62°C was used in view of the slightly lower Tm of one of the primers (C719). The expected product of approximately 1.3 kb was obtained when the VK5 plasmid was used as a template, but multiple non-specific bands were obtained using genomic DNA. Since the VK5/*Msp*I primers were working well, the VK5/*Kpn*I primers were not tested further.

5.3.4 PCR of ATPL

The sequence of the 3' untranslated region of the ATPL locus (Gillespie *et al.*, 1991) is shown in Appendix 1 with its complementary strand. This was used in the PRIMER (Version 0.5) program to design forward and reverse primers: these are shown on the sequence in Appendix 1. No amplification product was obtained with the first set of primers tested (B928 and B927), but the pairs C18 (forward) and C19 (reverse) proved successful. The primer sequences are as follows:-

C18 5' CCC GCC CCC AGT AGT TGG TC 3' (Tm = 68.1°C)

C19 5' AAC AGG AAC CCA CAA GGC GAC A 3' (Tm = 68.2°C)

These primers should give a product 295 bp in length, which contains a unique *Bgl*II site. An annealing temperature of 63°C was used, and PCR was carried out as described in Materials and Methods. Using the above conditions, an amplification

product of approximately 300 bp containing a *Bgl*II site at the expected position was obtained. The optimal $MgCl_2$ concentration was found to be 1.5 mM.

5.3.5 YAC library screening

CEPH YAC library

The CEPH YAC library was screened with primers to the GGG1 (D16S259), ATPL and VK5 (D16S94) loci. Screening was performed by PCR, and the product visualised on agarose gels by ethidium bromide staining. The results are as follows:-

1. GGG1 (D16S259) - PCR screening of the 113 master pools yielded a single positive product of the correct size (1 kb) in pool 18 (Figure 5.4). The corresponding 96-well plates (69, 70, 71, and 72) were sent from CEPH and screened by colony hybridisation. Two clearly positive colonies were seen in replicate, corresponding to clones 71B2 and 71B4.

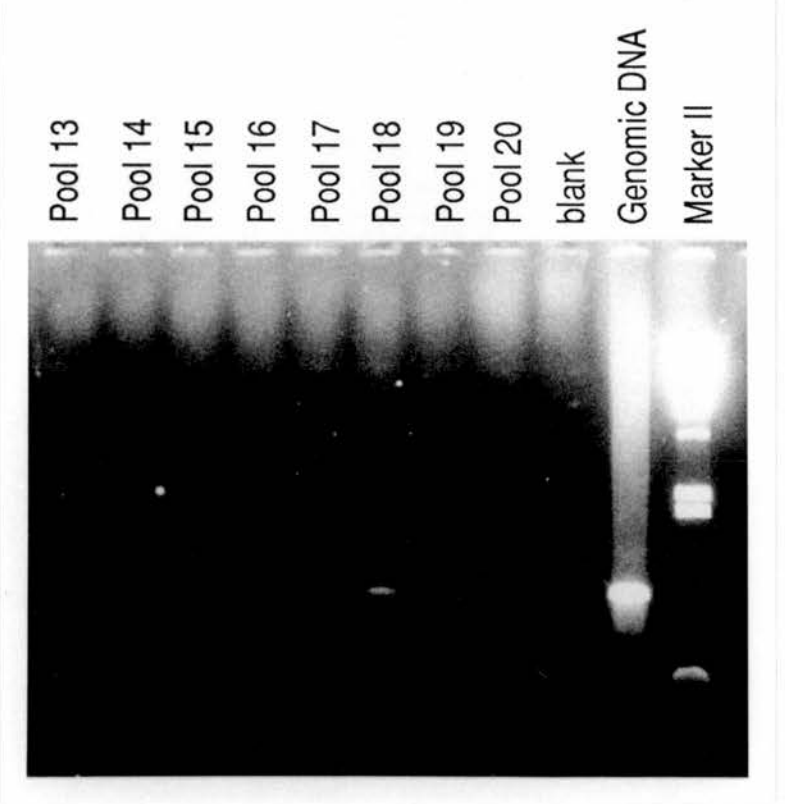
2. ATPL - PCR screening of the master pools revealed a single positive in pool 94 (Figure 5.5). Subpools, consisting of rows x columns, were obtained but no positive YAC clone could be identified after secondary screening. Secondary screening was carried out by PCR, followed by Southern transfer and probing of the filter with labelled PCR product from human genomic DNA.

3. VK5 (D16S94) - no positive master pools were identified using the primers VF1 and VR1.

ICI YAC library

The ICI YAC library was screened with primers to the GGG1 (D16S259), 26-6 (D16S125), SM33p22, ATPL and VK5 (D16S94) loci. Screening with primers to the GGG1 (D16S259) and 26-6 (D16S125) loci was initially performed by PCR alone,

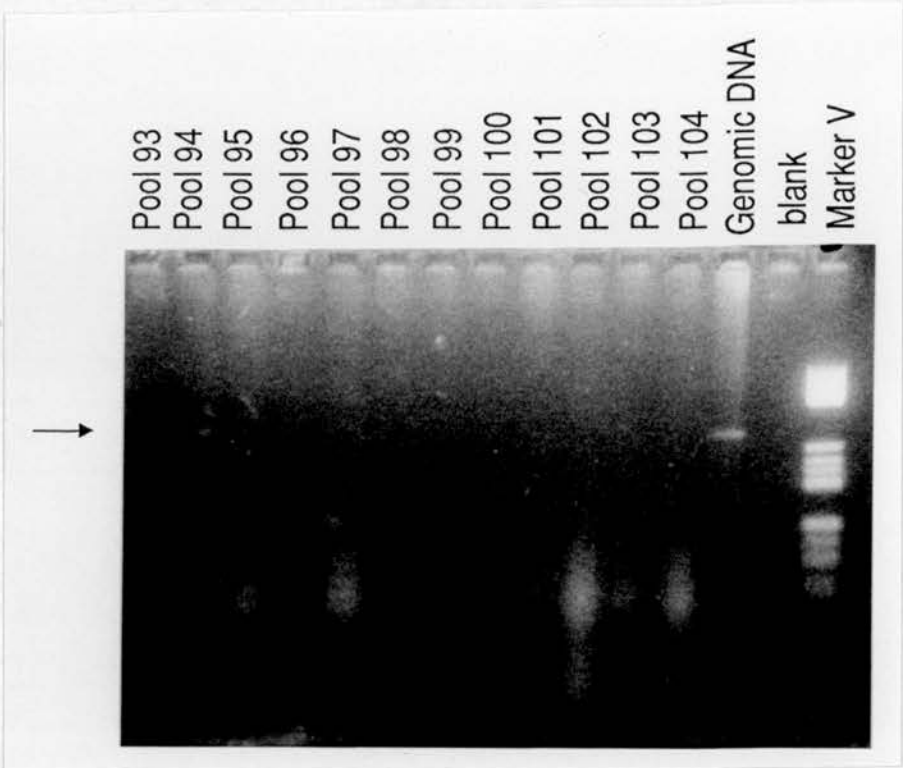
Figure 5.4



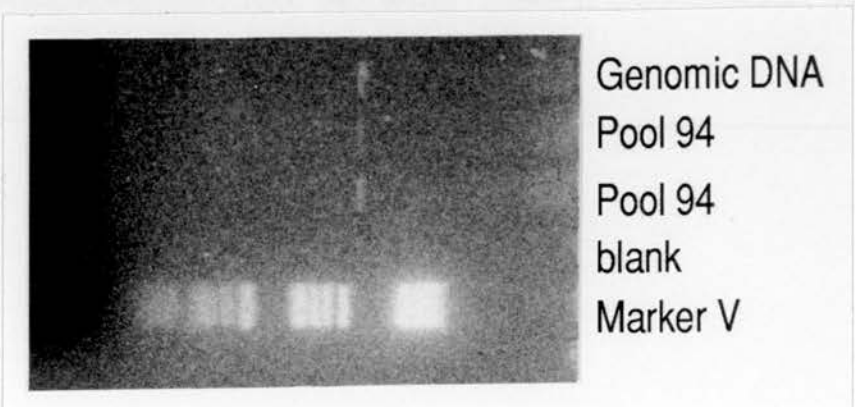
Screening of the CEPH YAC library with primers to the GGG1 (D16S259) locus yielded a product from master pool 18. This product was the same size as the control product obtained from genomic DNA. Clones 71B2 and 71B4 were isolated from the corresponding 96-well plates by colony hybridisation.

Figure 5.5

a)



b)



Screening of the CEPH YAC library with primers to the ATPL locus yielded a faint product from master pool 94 (Figure 5.5a). Re-screening of this pool confirmed that it was positive with the ATPL primers (Figure 5.5b). The product was the same size as the control product obtained from genomic DNA. No positive YAC clone was identified after secondary screening.

visualising any product on the gel by ethidium bromide staining. No positive pools were obtained, however, so the sensitivity was increased by Southern blotting the gels and probing the resultant filters. The results are as follows:-

1. GGG1 (D16S259) - no positive pools were obtained on screening by PCR with visualisation of the product by ethidium bromide staining.

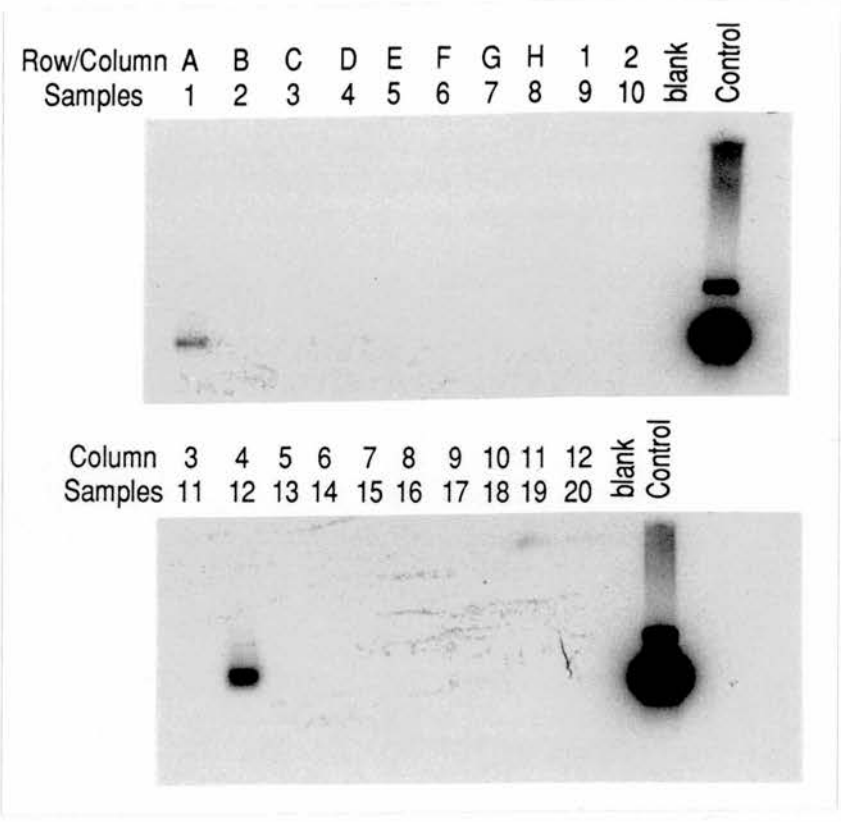
2. 26-6 (D16S125) - three positive pools (M14, M26 and M32) were obtained on screening the master pools. The M14, M26 and M32 subpools were therefore screened, and a single positive subpool, 32A, was identified. Tertiary screening of the 32A rows x columns revealed a clear amplification product of the correct size (430 bp) in tubes 1 and 12, corresponding to YAC clone 32A-A4 (Figure 5.6).

3. SM33p22 - master pool M14 was positive with primers to this locus, and screening of the subpools revealed 14C to be positive. Despite many attempts, however, no YAC clone could be isolated from the 14C rows x columns.

4. ATPL - three positive master pools (M25, M31 and M34) were obtained after primary screening. Secondary screening of the M25, M31 and M34 subpools found only subpool 31A to be positive (Figure 5.7). Tertiary screening of the 31A rows x columns revealed a weakly positive band in tube 8 and a strongly positive band in tube 18, corresponding to YAC clone 31A-H10.

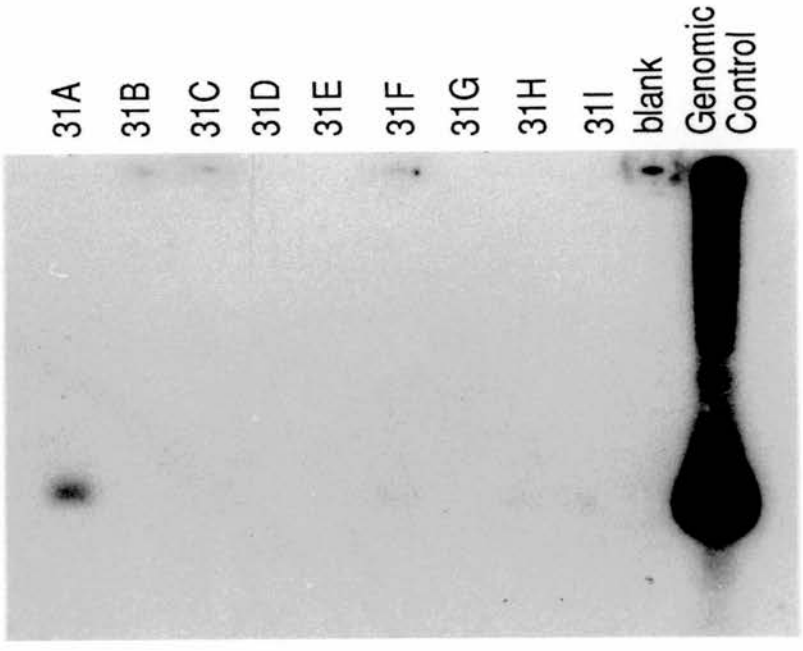
5. VK5 (D16S94) - three positive pools (M9, M15 and M32) were obtained on screening the master pools. Screening of the M9, M15 and M32 subpools revealed a single positive subpool, 32A. Tertiary screening of the 32A rows x columns revealed clear products, again corresponding to YAC clone 32A-A4 (Figure 5.8).

Figure 5.6



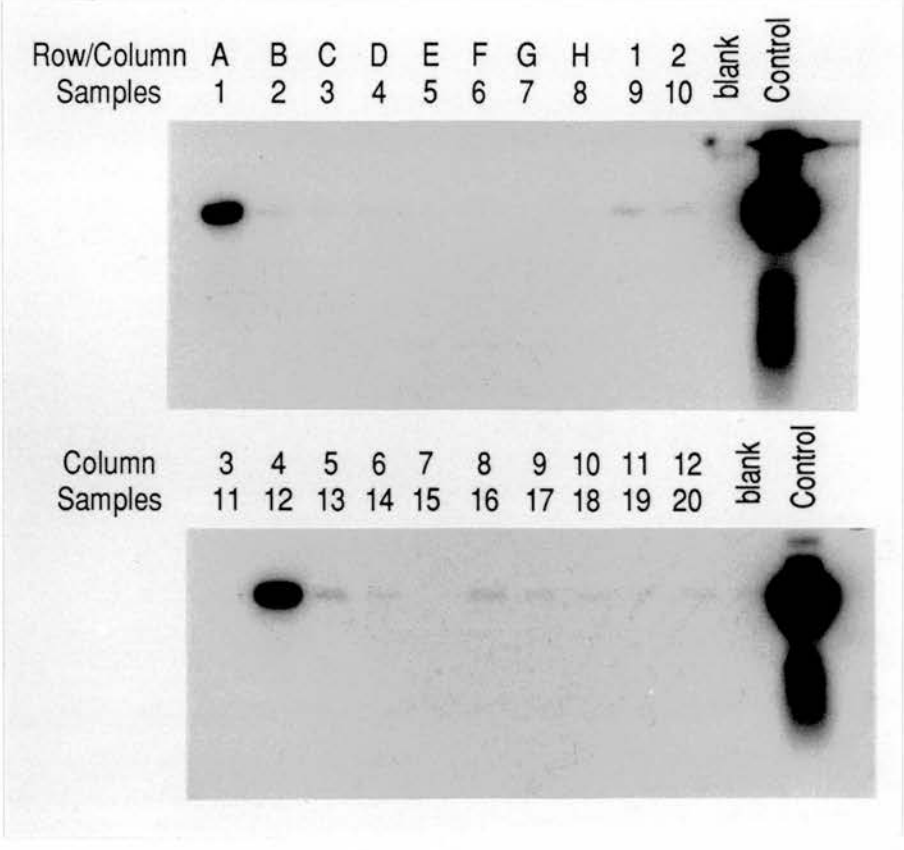
Autoradiograph showing tertiary screening of the ICI YAC library subpool 32A. The 32A rows x columns were screened using primers to the 26-6 (D16S125) locus. A clear product was seen in samples 1 and 12, which was of the same size obtained from genomic DNA. This isolates 32A-A4 as the positive clone.

Figure 5.7



Autoradiograph showing secondary screening of the ICI YAC library M31 subpools with the C18/C19 primers to the ATPL locus. A clear product of the correct size was seen in subpool 31A.

Figure 5.8

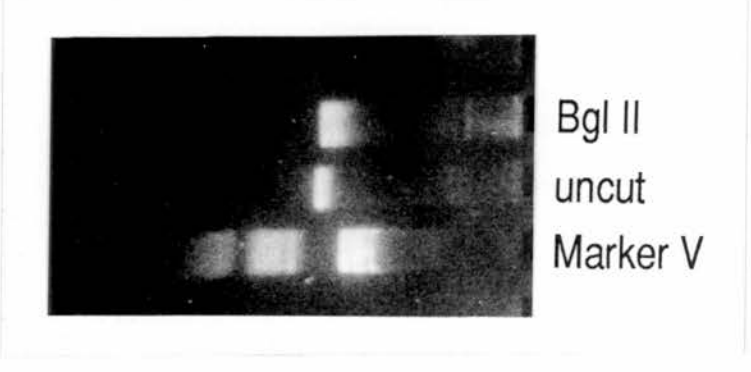


Autoradiograph showing tertiary screening of the 32A rows x columns with primers to the VK5 locus. Clear products of the correct size were obtained in samples 1 and 12, corresponding to YAC clone 32A-A4

St. Mary's Hospital Medical School mouse YAC library

Screening was performed by PCR followed by Southern blotting and probing. The annealing temperature of the PCR programme was reduced, initially by 3°C, to allow for species differences in sequence, and the primers tested on mouse (ThyB1) DNA. Once the correct product was obtained, the primers were used to screen the mouse YAC library. No product could be obtained from ThyB1 using the GGG1 (D16S259) primers, despite reducing the annealing temperature to 55°C (from 68°C) and increasing the $MgCl_2$ concentration to 2.5 mM (from 1.5 mM). A product was obtained using the ATPL primers by decreasing the annealing temperature to 57°C (from 63°C): the product was clearly visible on ethidium bromide staining and was of similar size as the product obtained from human genomic DNA, but without the *Bgl*II site (Figure 5.9).

Figure 5.9



Ethidium bromide-stained gel showing the PCR product obtained from mouse (ThyB1) DNA using the C18/C19 primers to the ATPL locus. There is no change in size of the product after incubation with *Bgl*II, indicating that the *Bgl*II site present in human genomic DNA is absent in ThyB1 DNA.

A product was obtained using the SM33p22, 26-6 (D16S125) and VK5 (D16S94) primers by decreasing the annealing temperature to 60°C (from 66°C, 65°C and

63°C respectively). In the case of VK5 (D16S94), the $MgCl_2$ concentration was increased to 1.5 mM from 1.0 mM. For all three loci, the products obtained were not visible by ethidium bromide staining alone, but were clearly seen if the DNA was transferred to a filter and probed with labelled product from human genomic DNA.

The mouse YAC library was therefore screened with primers to the ATPL, VK5 (D16S94), SM33p22 and 26-6 (D16S125) loci. The results are as follows:-

1. ATPL - screening of the complex pools revealed a clear product in pool C (Figure 5.10a), and screening of the simple pools revealed C4 to be positive (Figure 5.10b). Rows x columns were prepared from the corresponding 96-well plates, and tertiary screening identified C4-I12 as the positive YAC clone.

2. VK5 (D16S94) - complex pools B, C and E were found to be positive on primary screening, and simple pools B5 and E9 on secondary screening. Rows x columns were prepared from the corresponding microtitre plates, but no reproducible positives were identified on tertiary screening.

3. SM33p22 - complex pools E and F were found to be clearly positive on primary screening (Figure 5.11), but no clearly positive pools could be identified on secondary screening.

4. 26-6 (D16S125) - as with the above locus, complex pools E and F were found to be weakly positive on primary screening but, again, no clearly positive simple pools could then be identified.

Screening of the three libraries therefore yielded a total of five positive clones: 71B2 and 71B4 from the CEPH library positive for the GGG1 (D16S259) locus; 32A-A4 and 31A-H10 from the ICI library positive for the 26-6 (D16S125)/VK5 (D16S94) and ATPL loci respectively; and C4-I12 from the St. Mary's mouse library positive for the ATPL locus.

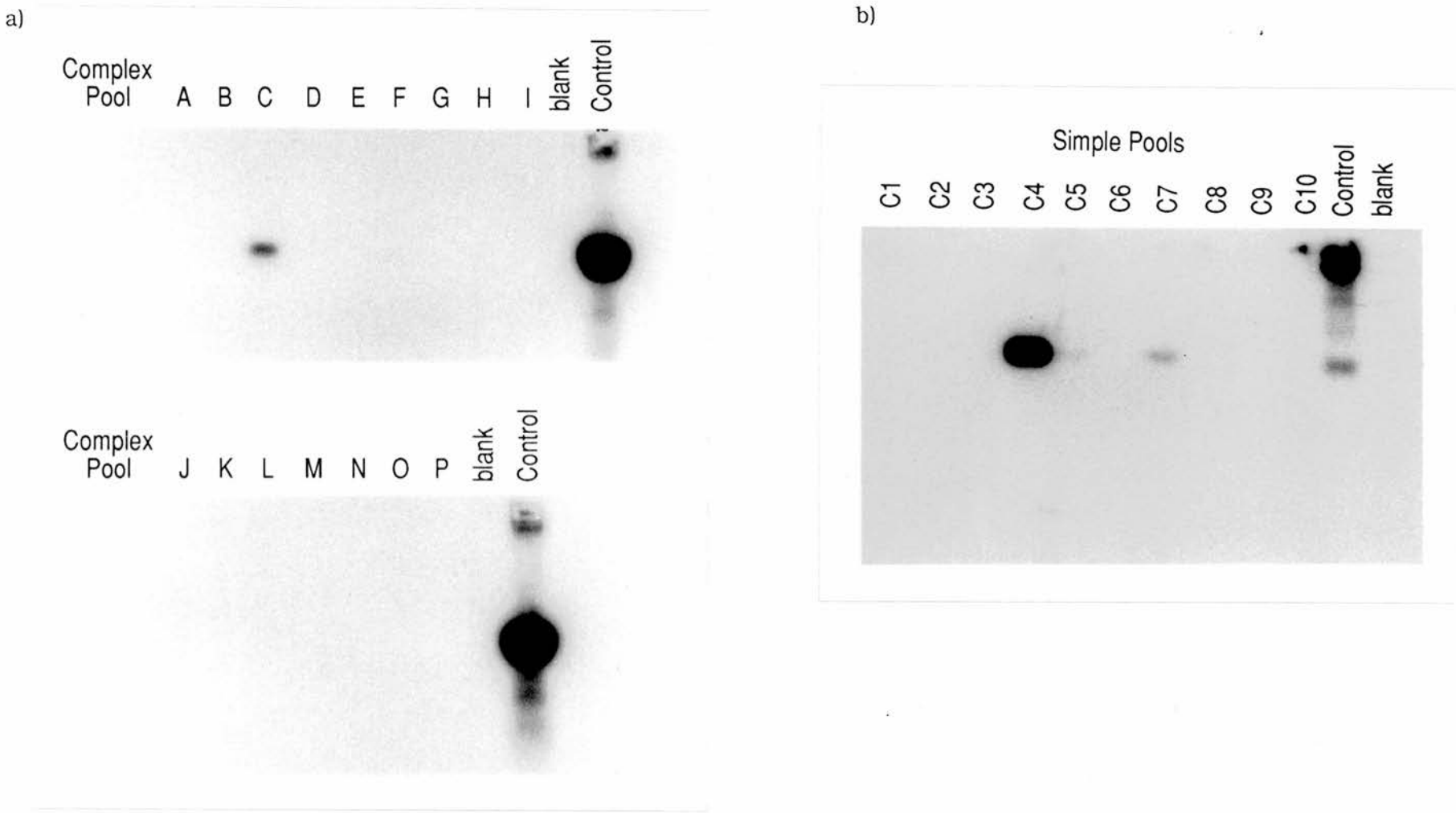
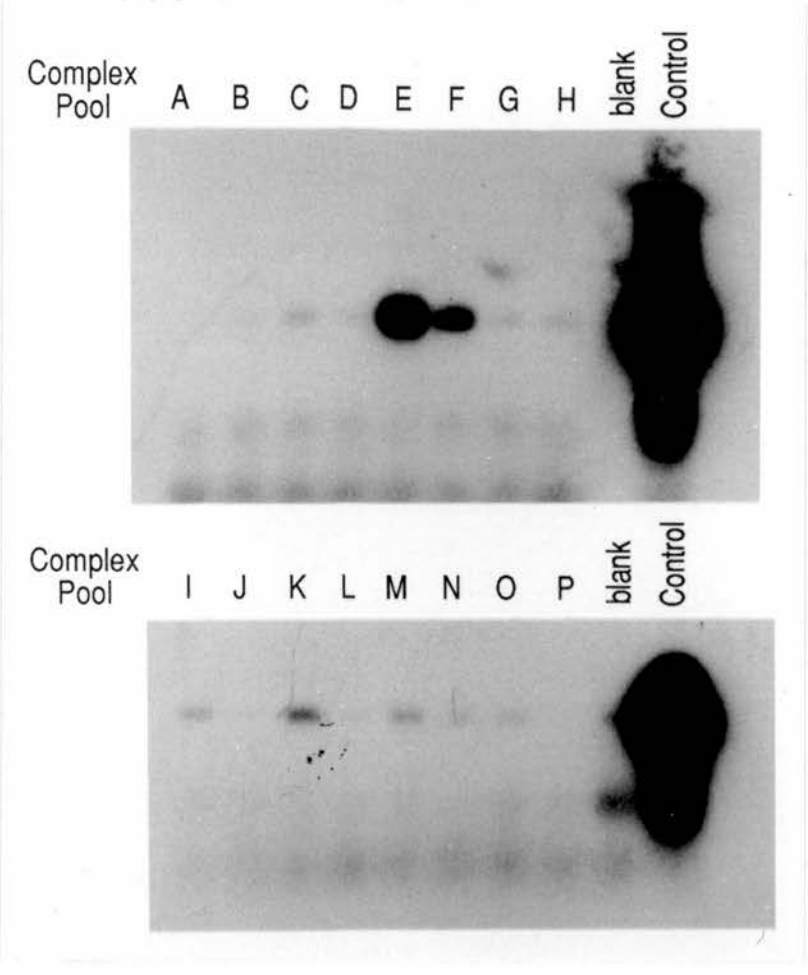


Figure 5.10 Autoradiographs showing the results of screening of the St. Mary's Hospital mouse YAC library with primers to the ATPL locus. Primary screening identified a product of the correct size in complex pool C (Figure 5.10a), and secondary screening showed simple pool C4 to be clearly positive (Figure 5.10b). Tertiary screening later identified YAC clone C4-I12 as positive for this locus.

Figure 5.11



Autoradiograph showing the results of primary screening of the St. Mary's Hospital mouse YAC library with primers to the SM3p22 locus. Complex pools E and F were found to be positive, with a product of the correct size demonstrated. Screening of the corresponding simple pools was unrewarding.

5.3.6 Characterisation of the YAC clones

71B2 and 71B4 (GGG1-positive human YAC)

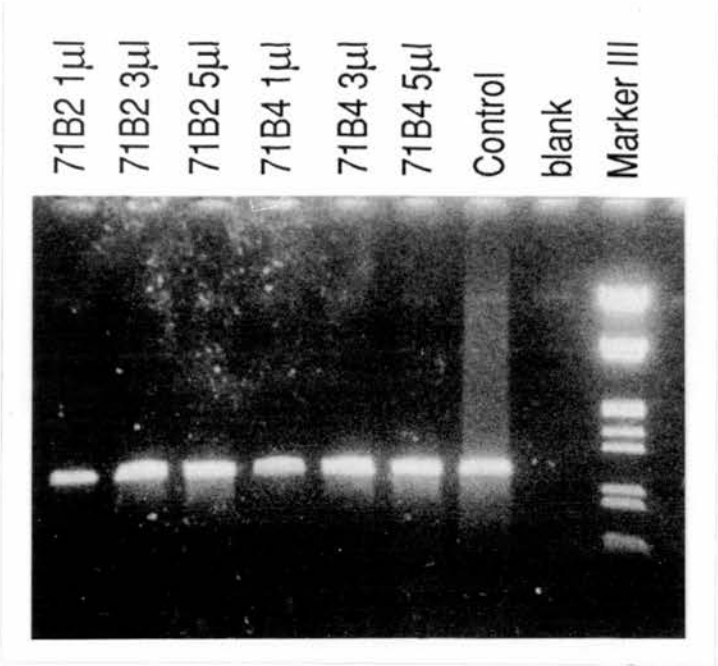
The clones were plated out onto Casamino agar, and plugs made from cultures of single colonies. Plug DNA was used as a template for PCR with primers to the GGG1 (D16S259), 26-6 (D16S125) and SM33p22 loci, and to the *Alu* primers TC-65 and 517 (Nelson *et al.*, 1989). A clear product of the correct size was obtained with both 71B2 and 71B4 using the GGG1 (D16S259) primers (Figure 5.12a), but no amplification product was obtained for either YAC using primers to 26-6 (D16S125) or SM33p22. PCR with the *Alu* primers gave the same band pattern for both YACs (Figure 5.12b) suggesting that they may be the same.

The yeast chromosomes were separated by pulsed field gel electrophoresis (PFGE), but no YAC band could be seen separate from the host yeast chromosomes on ethidium bromide staining (Figure 5.13). Following Southern blotting and probing of the filter with CMM65 (D16S84), which is located approximately 10 kb proximal to GGG1 (D16S259), a clear band was obtained for both 71B2 and 71B4 overlying the second smallest host yeast chromosome. 71B2 and 71B4 therefore both contain a YAC, 280 kb in size, with insert DNA from chromosome 16p positive for the GGG1 (D16S259) and CMM65 (D16S84) loci, but negative for the EKMDA (D16S83), 26-6 (D16S125) and SM33p22 loci.

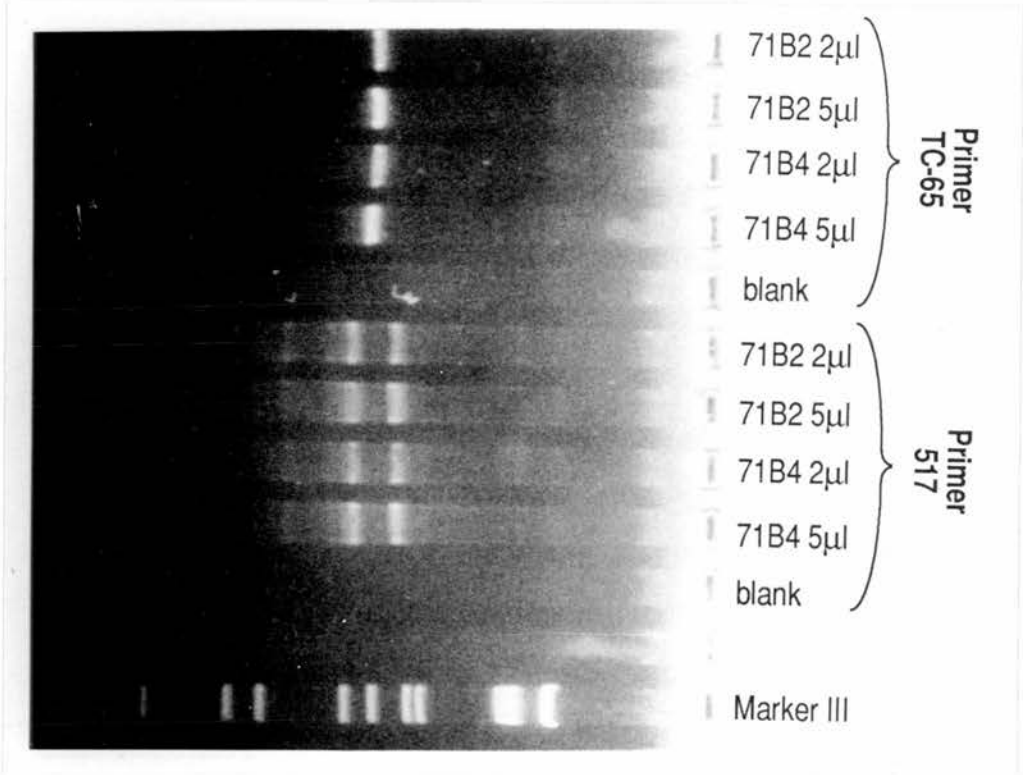
For each clone, the YAC band and its co-migrating host chromosome were cut out from LMP agarose, digested with *Bgl*II or *Rsa*I and ligated to catch-linkers. This material was amplified with 962L primers, labelled with biotin by nick translation and used for fluorescence *in situ* hybridisation. The *in situ* hybridisation was carried out by Dr. M. Breen. Both 71B2 and 71B4 gave a weak signal on 16p, but signals were also seen on six other chromosomes, suggesting either that both the clones were chimaeric, containing DNA from several different chromosomal locations, or that the clone contains low copy repetitive elements present on these other chromosomes.

Figure 5.12

a)

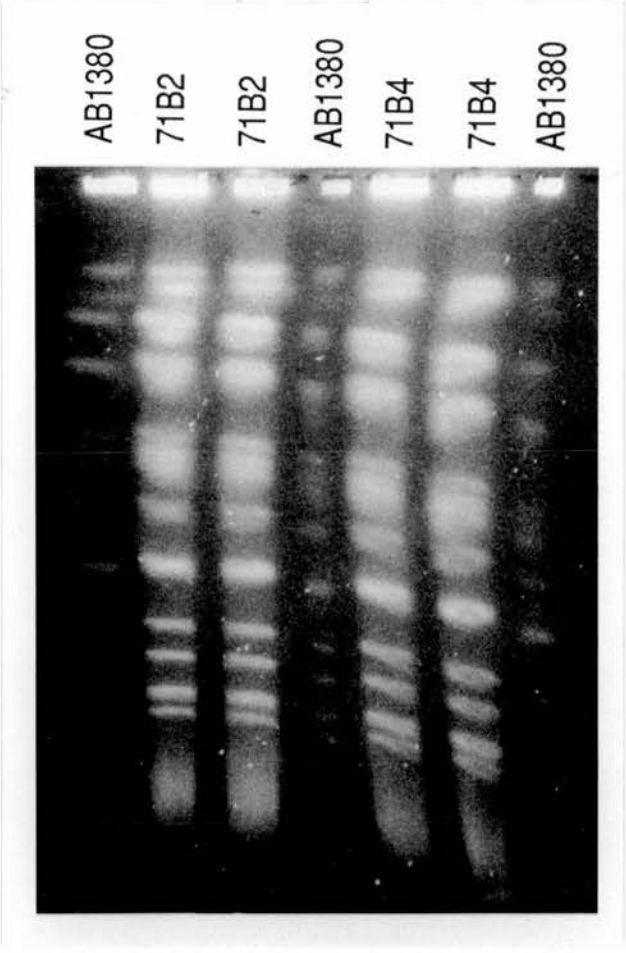


b)



a) PCR of 71B2 and 71B4 YAC plugs with primers to the GGG1 (D16S259) locus. As expected, a clear amplification product was seen in all the samples, indicating the presence of the GGG1 (D16S259) locus on both YACs.
b) PCR of 71B2 and 71B4 YAC plugs with the *Alu* primers TC-65 and 517. An identical band pattern was seen for both YACs, suggesting that they contain similar, or largely overlapping, DNA inserts.

Figure 5.13



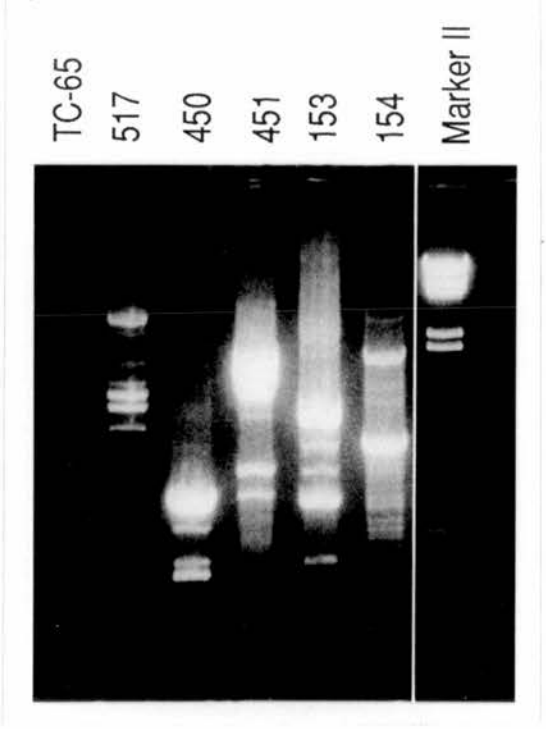
Ethidium bromide-stained pulsed field gel, showing separation of the yeast chromosomes in AB1380 and DNA from CEPH YAC clones 71B2 and 71B4. Separation was achieved using a pulse time of 70 sec for 21 hours, followed by a pulse time of 120 sec for 21 hours. The YACs cannot be clearly distinguished from the host yeast chromosomes. Probing of this gel with labelled GGG1 PCR product from genomic DNA indicated that the YAC was in a position overlying the second smallest yeast chromosome in both cases.

In conclusion, therefore, clones 71B2 and 71B4 contain the same potentially chimaeric YAC, with what may only be a small proportion of the insert DNA from 16p around the GGG1 (D16S259) locus.

32A-A4 (26-6/VK5-positive human YAC)

32A-A4 was grown up on Casamino agar and plugs made from cultures of single colonies. Plug DNA was used as a template for PCR with primers to the VK5 (D16S94), 26-6 (D16S125), ATPL and SM33p22 loci, and for inter-*Alu* PCR with primers TC-65, 517, 153, 154, 450 and 451. Clear amplification was seen following PCR with primers to VK5 (D16S94) and 26-6 (D16S125), but not with primers to ATPL or SM33p22. No bands were seen after PCR with TC-65, but all the other *Alu* primers amplified multiple bands (Figure 5.14).

Figure 5.14



PCR of 32A-A4 YAC plugs with *Alu* primers TC-65, 517, 450, 451, 153 and 154. No product was obtained using primer TC-65, but all the other primers gave multiple bands. This indicates that the 32A-A4 YAC contains several repeats of the *Alu* consensus sequence.

The yeast chromosomes were separated by PFGE and a YAC band was visible below the smallest host chromosome on ethidium bromide staining, approximately 230 kb in size (Figure 5.15). The DNA was transferred to a Hybond-N^+ filter by alkali blotting, and the position of the YAC confirmed by hybridisation with the probe 26-6 (D16S125). The filter was then stripped and probed with 218EP6 (D16S246) which is located proximal to VK5 (D16S94) (Figure 1.3). The YAC did not hybridise with this probe.

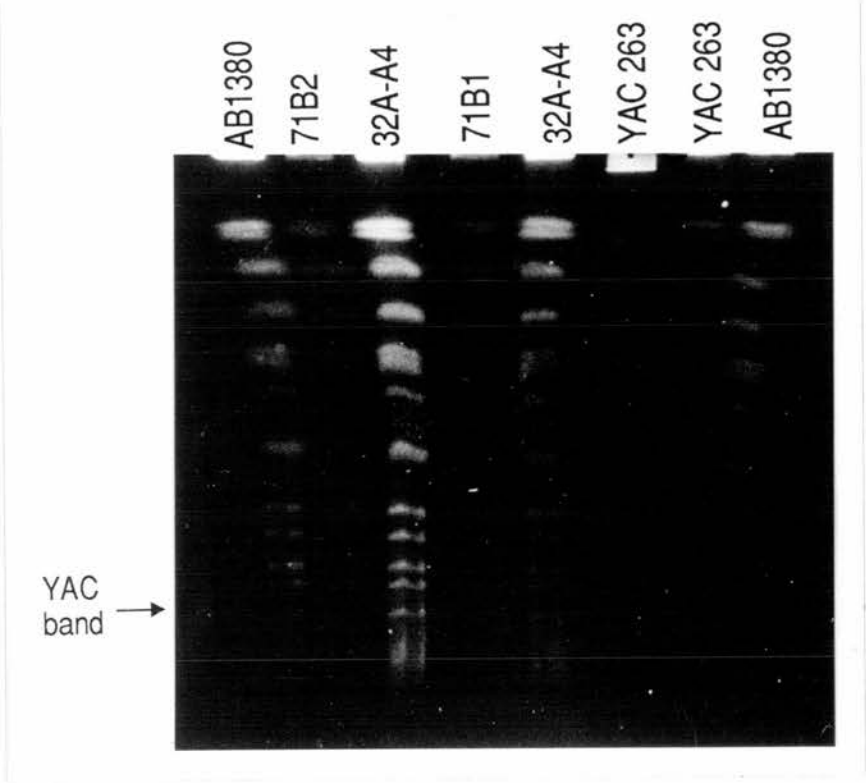
The YAC was cut out from LMP agarose (Figure 5.15b), digested with *Bgl*II or *Rsa*I, ligated to catch-linkers and amplified with 962L primers. The amplified material was labelled with biotin by nick translation and used for fluorescence *in situ* hybridisation. This *in situ* hybridisation was carried out by Ms. S. Boyle, and the YAC material was clearly shown to hybridise to 16p only. 32A-A4 is therefore a non-chimaeric 230 kb YAC with the VK5 (D16S94) and 26-6 PROX and/or DIS (D16S125) loci present. The YAC does not extend to include either the SM33p22 or 218EP6 (D16S246) loci.

Preliminary pulsed field mapping of the YAC was performed by digesting YAC plugs with the following rare cutter enzymes: *Not*I, *BssH*II, *Eag*I, *Ksp*I, *Sal*I, *Sfi*I, *Nru*I, and *Mlu*I. Half of each of the digested plugs were separated by PFGE, the DNA transferred to a filter and the filter probed sequentially with left and right arm vector probes. The restriction map showing the two ends of the YAC is shown in Figure 5.16, and shows the presence of at least two CpG-rich clusters which could represent CpG islands within the insert DNA.

31A-H10 (ATPL-positive human YAC)

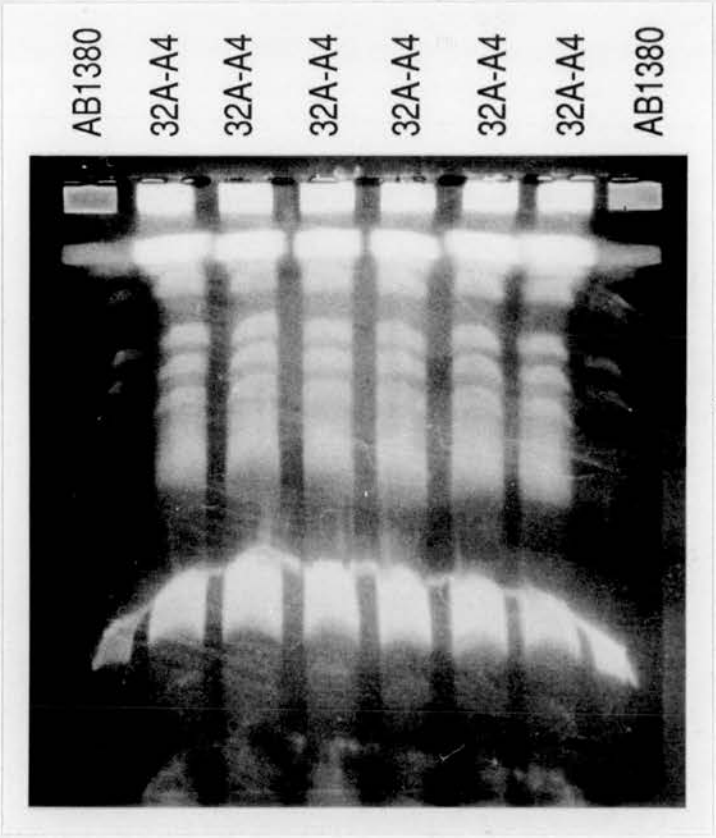
The clones were plated out onto Casamino agar, and plugs made from cultures of single colonies. Plug DNA was used as a template for PCR with primers to the ATPL and VK5 (D16S94) loci, and the *Alu* primers 517, 153, 154, 450 and 451. No amplification was obtained with the C18/C19 ATPL primer pair that had been used to pull out the YAC, or with the VK5 (D16S94) primers, but bands were obtained with the *Alu* primers suggesting the presence of human DNA in the YAC.

Figure 5.15a



Ethidium bromide-stained pulsed field gel, showing separation of the yeast chromosomes in AB1380, DNA from ICI YAC clone 32A-A4 and DNA from control YAC clones. Separation was achieved using a pulse time of 70 sec for 21 hours, followed by a pulse time of 120 sec for 21 hours. The YAC can be clearly distinguished below the smallest of the host yeast chromosomes in 32A-A4.

Figure 5.15b



Ethidium bromide-stained LMP pulsed field gel, showing separation of the yeast chromosomes in AB1380 and DNA from ICI YAC clone 32A-A4. The YAC band is seen below the smallest yeast chromosome. The YAC DNA was excised, digested and ligated to catch-linkers (see text).

Figure 5.16

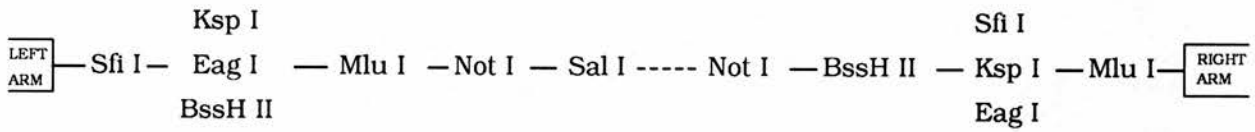


Diagram showing the position of restriction sites at each end of YAC 32A-A4. The map is not to scale. This preliminary restriction map was constructed from complete digests of 32A-A4 YAC plugs separated on a pulsed field gel and probed sequentially with left and right arm vector probes. Two CpG-rich clusters are seen, one close to each vector arm.

The yeast chromosomes were separated by PFGE, and a YAC band could be seen on ethidium bromide staining which was separate from, and smaller than, the host yeast chromosomes, approximately 180 kb in size. Following Southern transfer, the filter was probed with labelled ATPL PCR product, and a weak signal was obtained in a position consistent with the band previously seen. The filter was stripped and re-probed using pBR322 DNA, and a strongly positive band was seen in the same position confirming the presence of YAC vector sequence.

PCR was then performed on material taken directly from half each of nine YAC colonies, and amplification was obtained using the ATPL primers. Cultures were grown using -URA,-TRP selection from the other halves of the colonies and DNA plugs made. However, multiple attempts at PCR amplification of these plugs with the ATPL primers failed.

It therefore seems that clone 31A-H10 contains a YAC with a human DNA insert which is positive for the ATPL primer sequences when amplified from a colony, but if the clone is grown up and made into plugs this sequence is lost. This unstable YAC could not be isolated from a pulsed field gel with any degree of confidence, and was therefore not analysed further.

C4-I12 (ATPL-positive mouse YAC)

Clone C4-I12 was picked from the microtitre plate, grown up on Casamino agar and plugs made from cultures of single colonies. Plug DNA was used as a template for PCR with primers to the ATPL, 26-6 (D16S125) and VK5 (D16S94) loci. Clear amplification was seen with the ATPL primers, but not with primers to the other two loci. The yeast chromosomes were separated by PFGE and a possible YAC band was visible well below the smallest host chromosome on ethidium bromide staining (Figure 5.17a). The DNA was transferred to a Hybond-N^+ filter by alkali blotting, and the filter probed with the ATPL PCR product from murine genomic DNA template. A strongly positive band was seen at the base of the gel (Figure 5.17b), corresponding to a YAC of approximately 160 kb in size. Confirmation of the

Figure 5.17

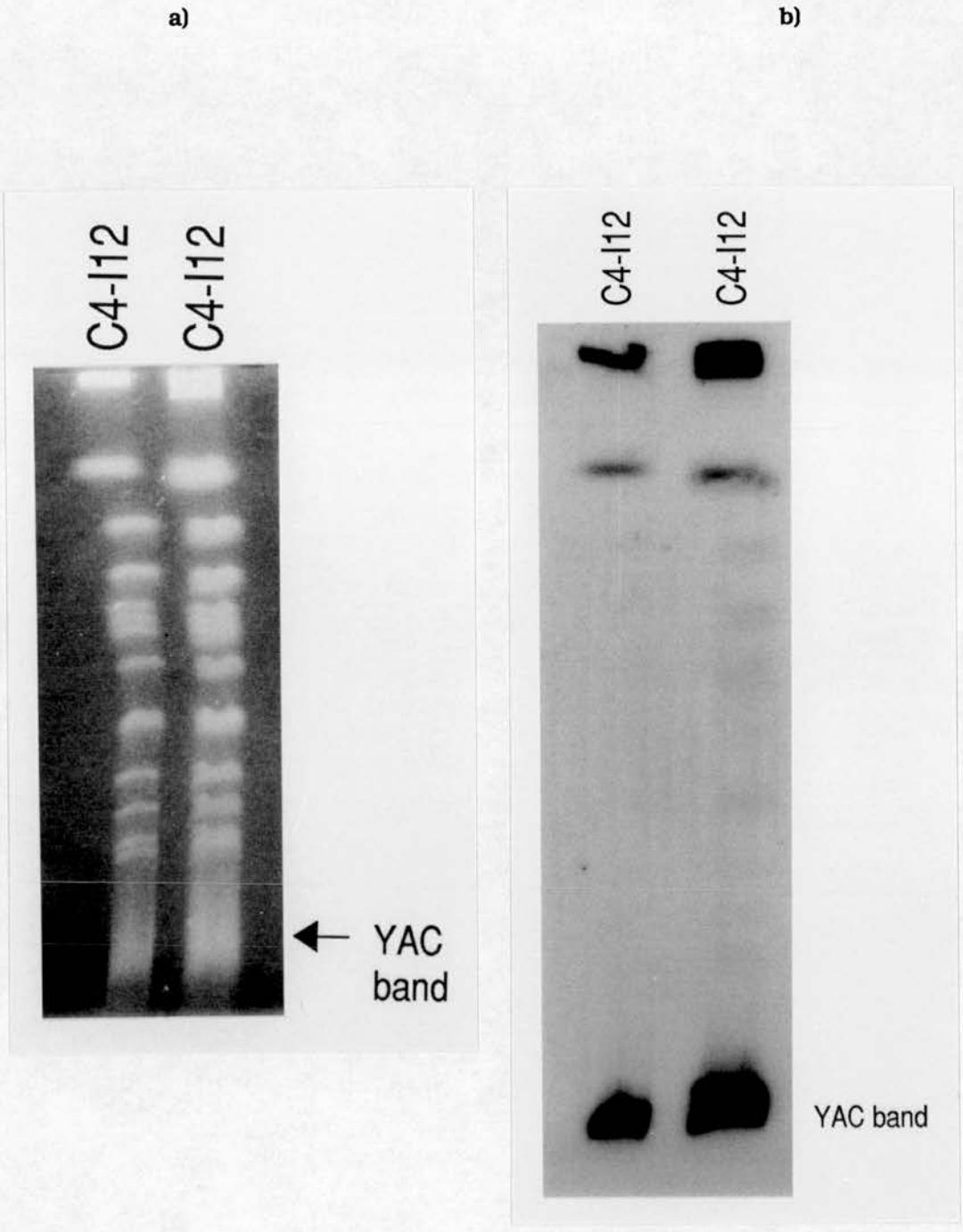
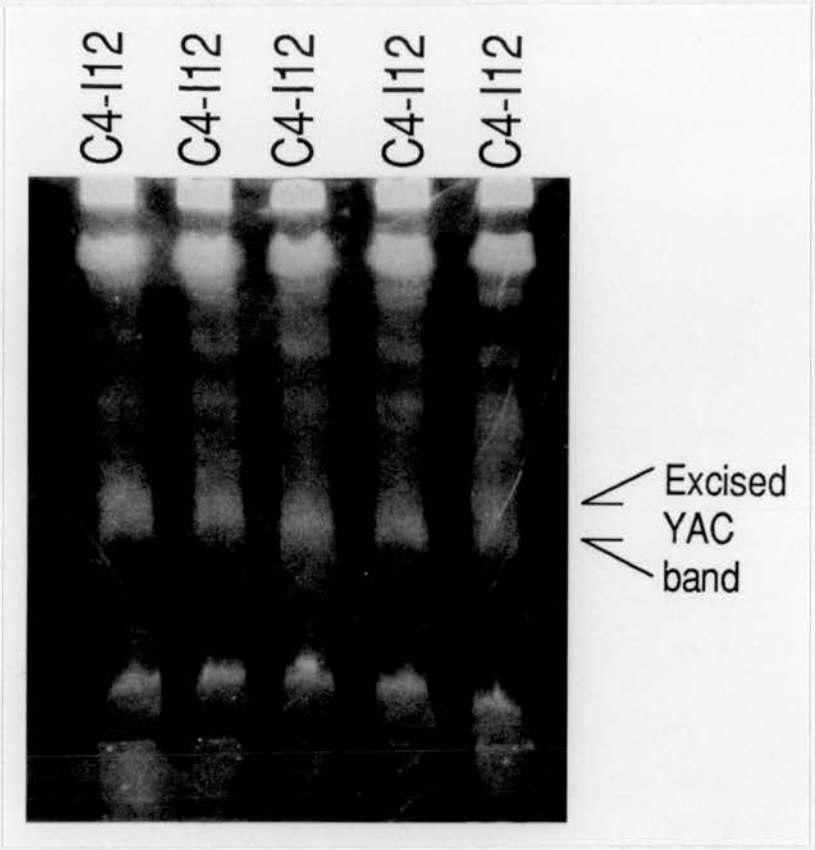


Figure 5.17a shows an ethidium bromide-stained pulsed field gel of YAC clone C4-I12 with a diffuse YAC band at the bottom of the gel, smaller than the smallest host yeast chromosome. Figure 5.17b is the autoradiograph of the filter corresponding to the gel in Figure 5.17a which has been probed with labelled pBR322. pBR322 recognises the vector sequence in the YAC, and confirms the position of the C4-I12 YAC band on the gel.

Figure 5.17

c)



d)

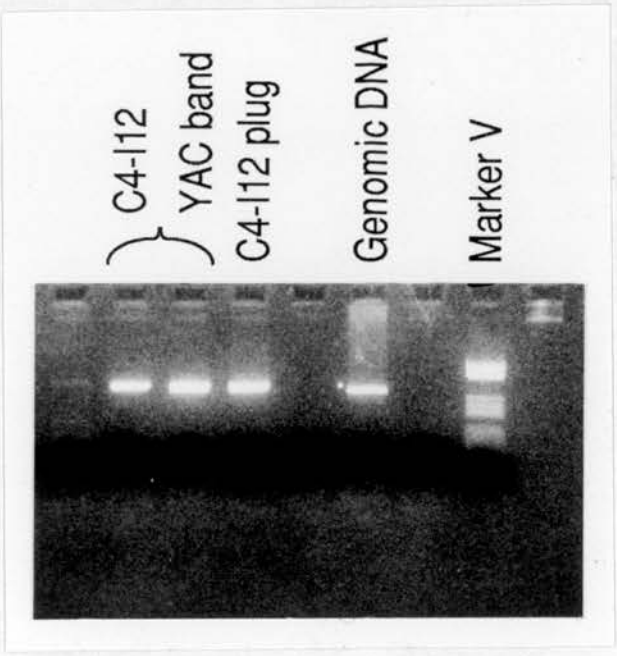


Figure 5.17c is the ethidium bromide-stained LMP pulsed-field gel from which the diffuse C4-I12 YAC band was excised. Figure 5.17d shows the PCR products from the C4-I12 plug and the excised YAC band using the ATPL primers. This confirms the presence of the ATPL sequence on the excised band.

presence of a YAC was obtained by stripping the filter and re-probing with pBR322, which again gave a strongly positive band in the same position.

The YAC was excised from LMP agarose (Figure 5.17c), and this material was used as a template for PCR using the ATPL primers. Good amplification was obtained, confirming the presence of the ATPL sequence in the excised YAC band (Figure 5.17d). The YAC DNA was then digested and ligated to catch-linkers. This material was PCR amplified using the 962L primers, labelled with biotin by nick translation and attempts made at fluorescence *in situ* hybridisation. This *in situ* hybridisation was carried out by Mrs. M. Harris. A clear signal was obtained on a single mouse chromosome using this material, but it was not possible to identify the chromosome.

Preliminary pulsed field mapping of the YAC was performed by digesting YAC plugs with the following rare cutter enzymes: *Not*I, *Eag*I, *Ksp*I, *Sal*I, *Sfi*I, *Nru*I, and *Mlu*I. Half of each of the digested plugs were separated by PFGE, the DNA was transferred to a filter and the filter probed sequentially with left and right arm vector probes. The restriction map showing the two ends of the YAC is shown in Figure 5.18, and shows the presence of at least one CpG-rich cluster within the YAC DNA insert.

Figure 5.18

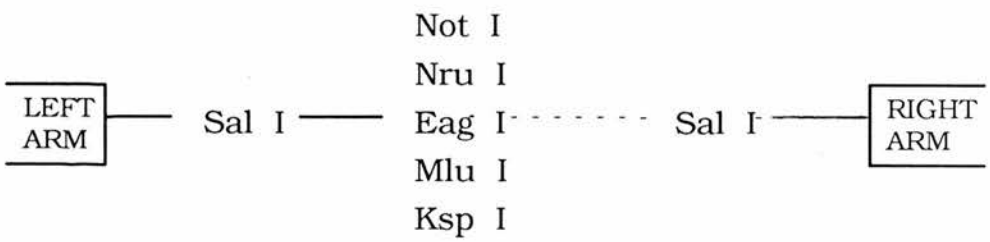


Diagram showing the position of restriction sites at each end of YAC C4-I12. The map is not to scale. This preliminary restriction map was constructed from complete digests of C4-I12 YAC plugs separated on a pulsed field gel and probed sequentially with left and right arm vector probes. A single CpG-rich cluster is seen, close to the left arm.

5.3.7 YAC-cDNA hybridisation

The details of the method used are described in Materials and Methods.

Biotin-streptavidin binding

A trial run was performed in order to ensure that the Geneclean™ procedure, used to remove unused biotinylated primers from the PCR reaction, did not affect binding of the biotinylated product to the Dynabeads™. 1 µl of a plasmid containing a catch-linkered insert was used as the template for PCR. Samples were set up using either 962L or C232 (biotinylated 962L) as the primer and PCR was carried out. 100 µl of each product was then either extracted with chloroform and using the Geneclean procedure, or extracted only with chloroform. 25 µl Dynabeads was added to each tube and treated as in the protocol. Both the Geneclean-ed and the non-Geneclean-ed biotinylated C232 products bound to the Dynabeads, with no detectable DNA remaining in the supernatant (Figure 5.19). The Geneclean procedure had no effect on the binding to Dynabeads.

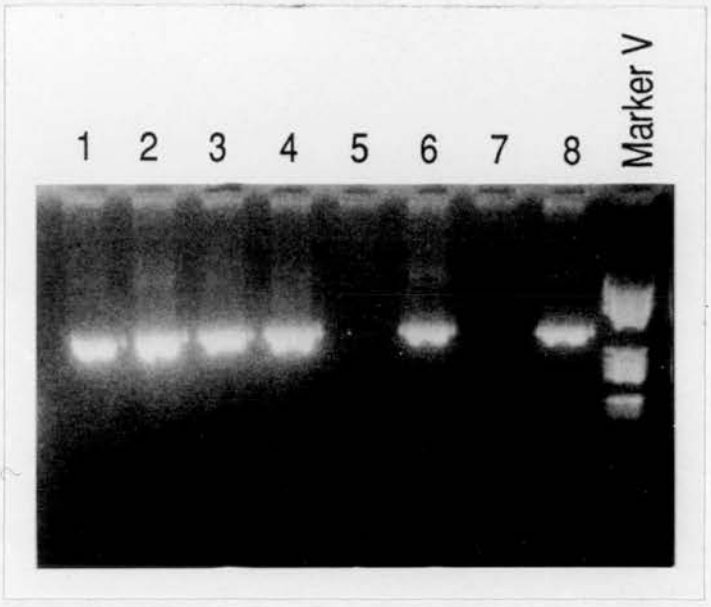
Titration of cDNA libraries

An adult and a foetal kidney cDNA library were available for screening. The titre of the adult library was found to be 2.35×10^{10} pfu/ml. The titre of the foetal library was found to be 1.25×10^{11} pfu/ml.

Amplification of cDNA

Primary PCR amplification of the cDNA library was carried out using the λgt10 primers. Amplification of a good size range of inserts was obtained for both libraries with an extension time of three minutes (Figure 5.20a). Further amplification of this material using the same PCR programme led to a significant reduction in the size of the amplified inserts. This was overcome by increasing the extension time to 7 minutes per cycle.

Figure 5.19



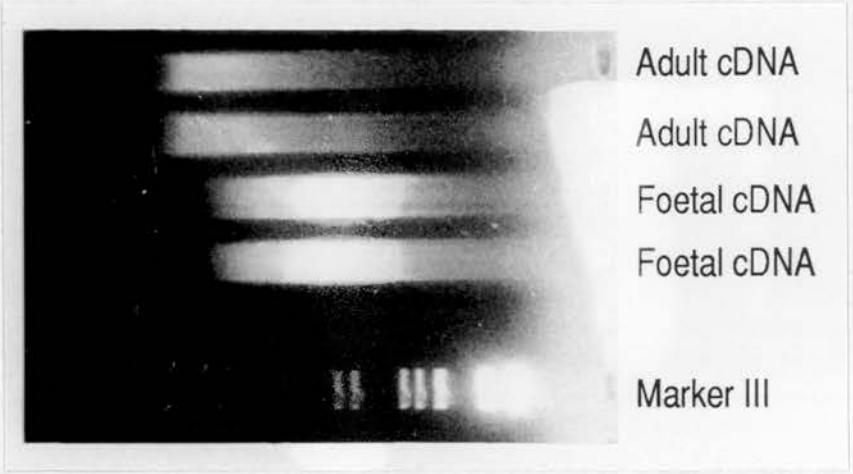
DNA was PCR amplified using either the 962L (non-biotinylated) primer or the C232 (biotinylated) primer. In samples 962LA and C232A, the PCR product was extracted with chloroform and purified using the Geneclean procedure. In samples 962LB and C232B, the PCR product was extracted with chloroform and not further purified. Each of the four samples was incubated with Dynabeads as described in Materials and Methods. The samples on the gel are as follows:-

1. 962LA added to Dynabeads and the supernatant removed immediately after incubation
2. 962LA prior to the addition of Dynabeads
3. 962LB added to Dynabeads and the supernatant removed immediately after incubation
4. 962LB prior to the addition of Dynabeads
5. C232A added to Dynabeads and the supernatant removed immediately after incubation
6. C232A prior to the addition of Dynabeads
7. C232B added to Dynabeads and the supernatant removed immediately after incubation
8. C232B prior to the addition of Dynabeads

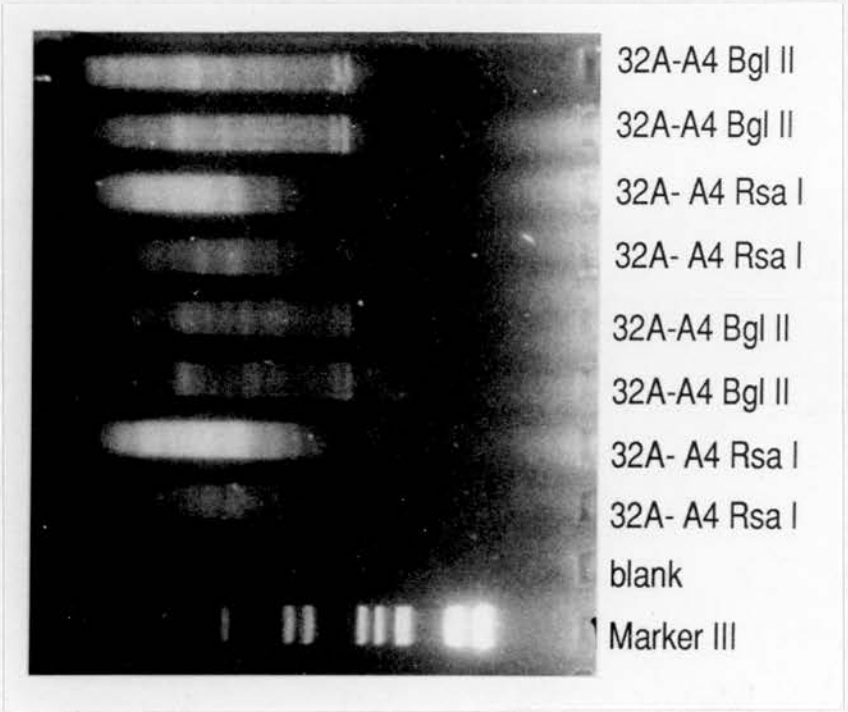
DNA is therefore present in all samples prior to the addition of Dynabeads. Samples 962A and 962B are non-biotinylated, and there was no/little binding of the sample to the Dynabeads with a significant amount present in the supernatant. Samples C232A and C232B are biotinylated, and there was complete binding of both samples to the Dynabeads with none being present in the supernatant. Purification with Geneclean has no measurable effect on this binding.

Figure 5.20

a)



b)



a) Ethidium bromide-stained gel showing PCR amplification of adult and foetal kidney cDNA libraries using the λgt10 forward and reverse primers.
b) Ethidium bromide-stained gel showing PCR amplification of YAC 32A-A4 that has been digested with either *Bgl*II or *Rsa*I and ligated to catch-linkers as described in Materials and Methods. PCR was performed using the biotinylated C232 primers and either 1μl of catch-linkered material (upper 4 lanes) or 1 μl 1:10 diluted material (lower 4 lanes).

Amplification of 32A-A4 YAC DNA

The 32A-A4 YAC DNA was amplified using biotinylated 962L primers (C232). Good amplification was obtained using the same PCR programme as for the non-biotinylated primers (Figure 5.20b).

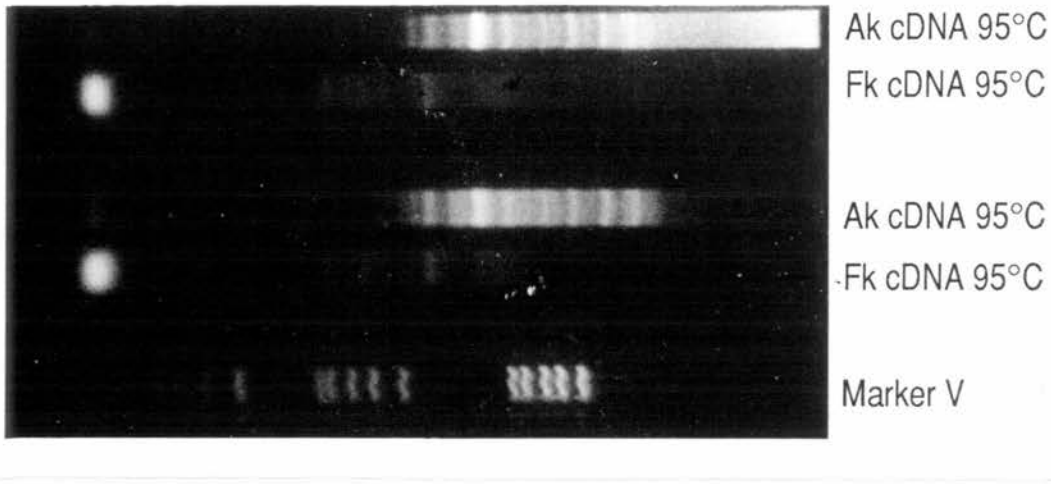
YAC-cDNA hybridisation

Hybridisation of biotinylated YAC DNA and cDNA inserts, both of which had been competed with *Cot*I DNA, was performed as described in Materials and Methods. 1 µl of material eluted from Dynabeads at 75°C, 85°C and 95°C after the first round of hybridisation was amplified with the λgt10 primers, using an extension time of 7 min for 25 cycles. A good yield was obtained with the adult library, and a lower but visible yield for the foetal library. Within the background smear, however, clear bands could be seen. The adult cDNA material eluted at 95°C, and the foetal cDNA material eluted at 85°C and 95°C, were therefore amplified in sufficient quantity to allow a second round of hybridisation to be performed. 1 µl of material eluted at 95°C from the second round of hybridisation was amplified with the λgt10 primers, using an extension time of 7 min for 20 cycles; 5 µl of product was then placed in fresh PCR mix and taken through a further 10 cycles of PCR. Multiple clear bands were seen, especially in the adult library (Figure 5.21a). Each band or "fraction" was then amplified from a needle stab of the gel. Clear bands were obtained for the smaller fractions, but the larger fractions failed to amplify despite increasing the extension time to 10 minutes per cycle (Figure 5.21b).

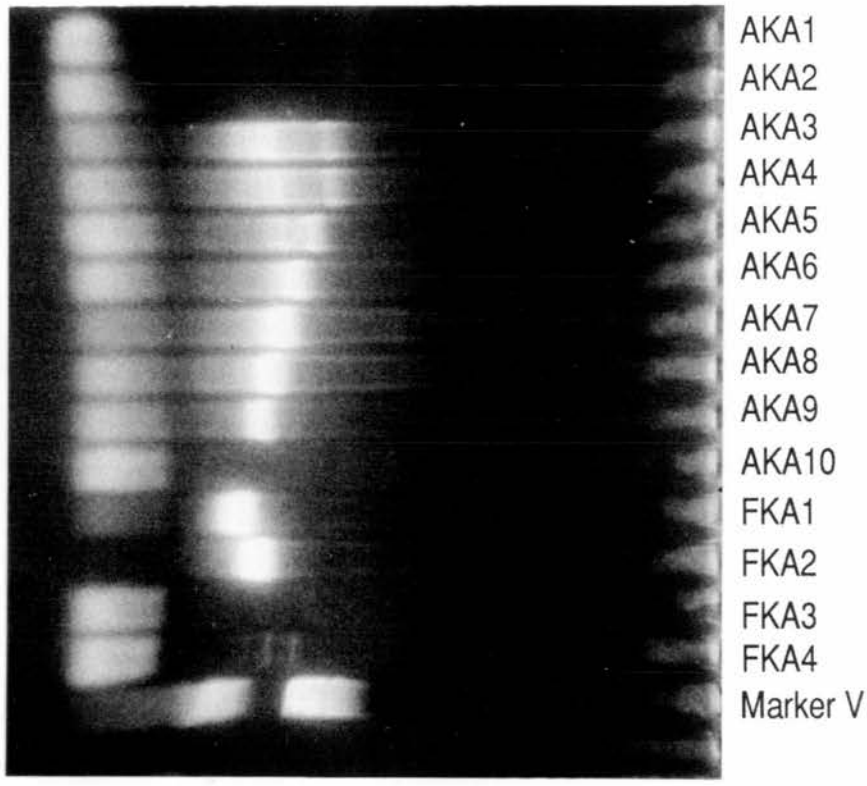
Two of these fractions (A7 and A8 from the adult library) were selected at random, re-amplified by PCR, and the blunt-ended fragments subcloned into pBluescript. DNA was directly amplified from recombinant colonies, using M13 primers, and inserts of differing sizes were obtained (Figure 5.22a). 25 µl of product from the A7 colonies were digested with *Eco*RI, and 4 different sizes or orientations of inserts were seen (Figure 5.22b). As expected, therefore, the fraction consists of a number of different hybrid products within a very narrow size range. Since a number of fractions with a number of hybrids have been obtained, further

Figure 5.21

a)



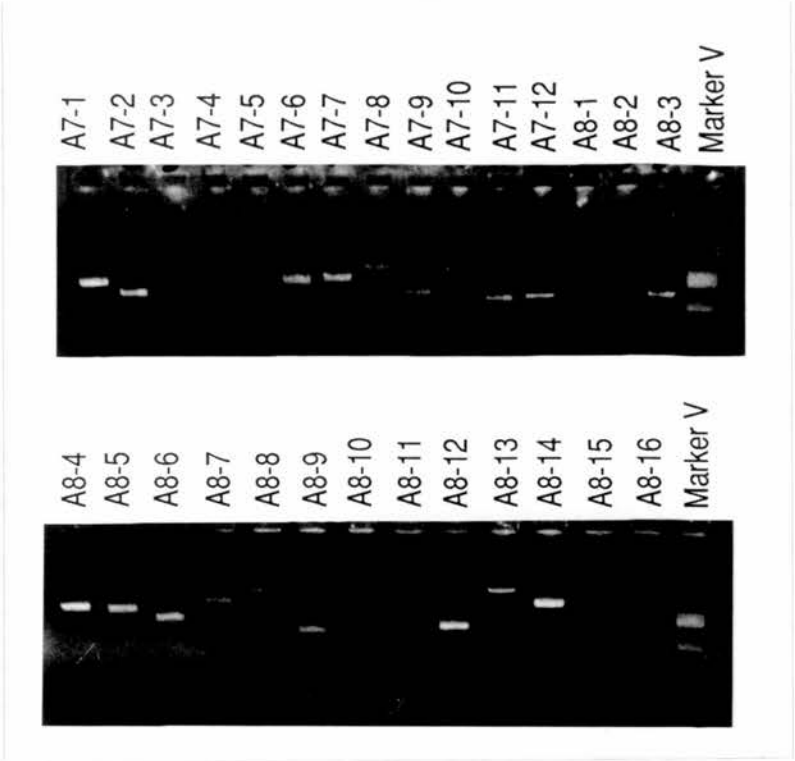
b)



a) 1 µl of material eluted at 95°C from the second round of YAC-cDNA hybridisation was amplified with the λgt10 primers as described in the text. Better yields were obtained for the adult library than the foetal library, but multiple bands could be seen in each sample. The samples in the lower two lanes have been through the 'clean-up' programme.

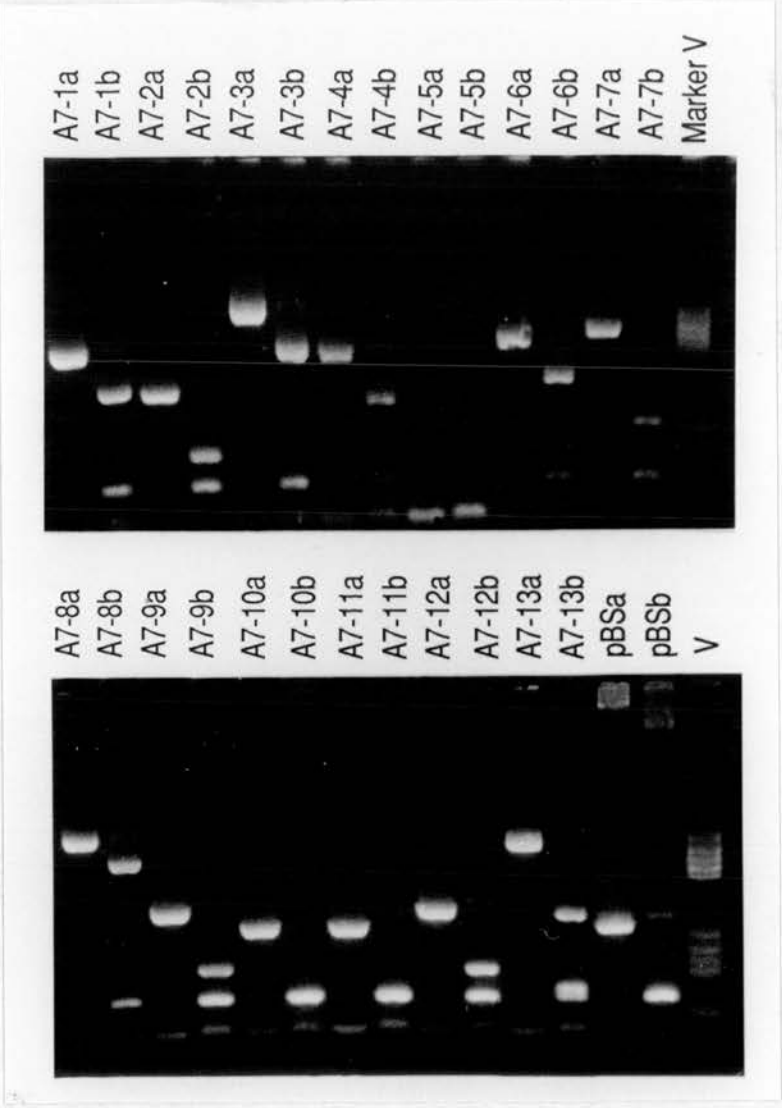
b) A needle stab was taken from each band of the above gel and PCR amplification performed using the λgt10 primers. Good amplification of the smaller bands was obtained, but the larger bands have failed to amplify.

Figure 5.22a



Fractions AKA7 and AKA8 were cloned into pBluescript and DNA was directly amplified from randomly picked recombinant colonies using the M13 forward and reverse primers. Colonies A7-3, A7-4, A7-5, A7-10, A8-1, A8-2, A8-10 and A8-16 contain no insert. Within each fraction (AKA7 and AKA8) inserts of differing sizes have been cloned, and each fraction therefore consists of a number of different hybrid products within a very narrow size range.

Figure 5.22b



25 µl of PCR products from the AKA7 colonies were digested with *Eco*RI. The samples are run on the gel both uncut (A7-1a - A713a) and after digestion with *Eco*RI (A7-1b - A7-13b). Four different sizes or orientation of insert are seen: clones 1, 4 and 6; clones 2, 9, and 12; clones 3 and 8; and clone 13. Clones 10 and 11 behave as empty vector and clone 5 has not amplified.

experiments were carried out in order to try and reduce the number of fractions to be more closely characterised.

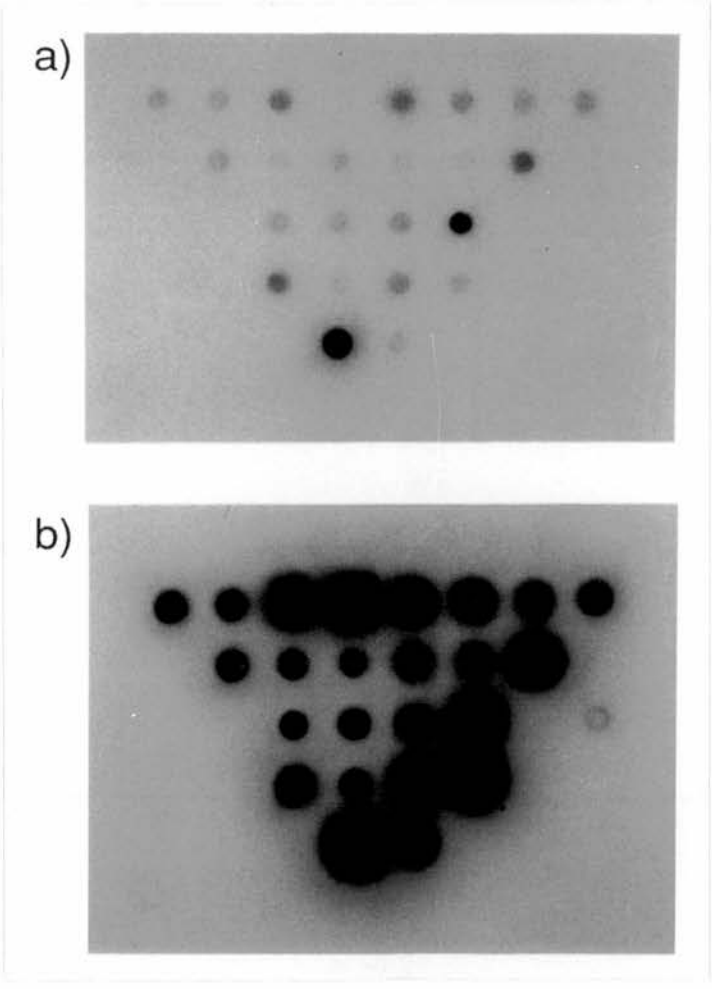
Assessment of fractions

1. Dot blot filters of the cDNA-YAC hybrid fractions were prepared and hybridised with labelled YAC DNA or with a labelled 2 kb *Alu* fragment (not including vector). All the fractions hybridised to both the YAC and the *Alu* sequence (Figures 5.23a and 5.23b). Both filters were washed to high stringency (0.1 x SSC/0.1% SDS/65°C). This would suggest that the cDNA hybrids do have sequence homology with the YAC used in their selection, but that these sequences also contain a high proportion of *Alu* repetitive sequences.

2. Agarose plugs containing AB1380, CHI-16 (a chromosome 16-only hybrid cell line), 32A-A4 or C4-I12 were digested with *Hind*III and electrophoresed in a 0.8% TBE agarose gel. The DNA was transferred to Hybond-N^+ filters and probed with labelled 26-6 (D16S125). This probe hybridised clearly to a single fragment of the digested YAC 32A-A4, but not to AB1380, CHI-16 or C4-I12, confirming that this sequence is present in single copy in the 32A-A4 YAC. The filter was then stripped and re-probed; first with the YAC-cDNA hybrid fraction AKA7 after *Cot*I competition, and then with the labelled *Alu* fragment. Both the YAC-cDNA hybrid and the *Alu* probe hybridised to several discrete bands within the *Hind*III digested YAC, but at least one band hybridised only with the cDNA and not with the *Alu* probe.

Since the *Alu* probe hybridised to several fragments in the digested YAC, the YAC must contain at least several *Alu* repetitive sequences. This is in keeping with the finding that PCR of 32A-A4 YAC DNA using the *Alu* primers led to the amplification of multiple bands. The AKA7 fraction also hybridised to several fragments in the digested YAC, nearly all of which were the same fragments as those hybridising to the *Alu* probe. This would suggest that AKA7 contains a high proportion of repetitive DNA, but that some single-copy sequence is also present.

Figure 5.23



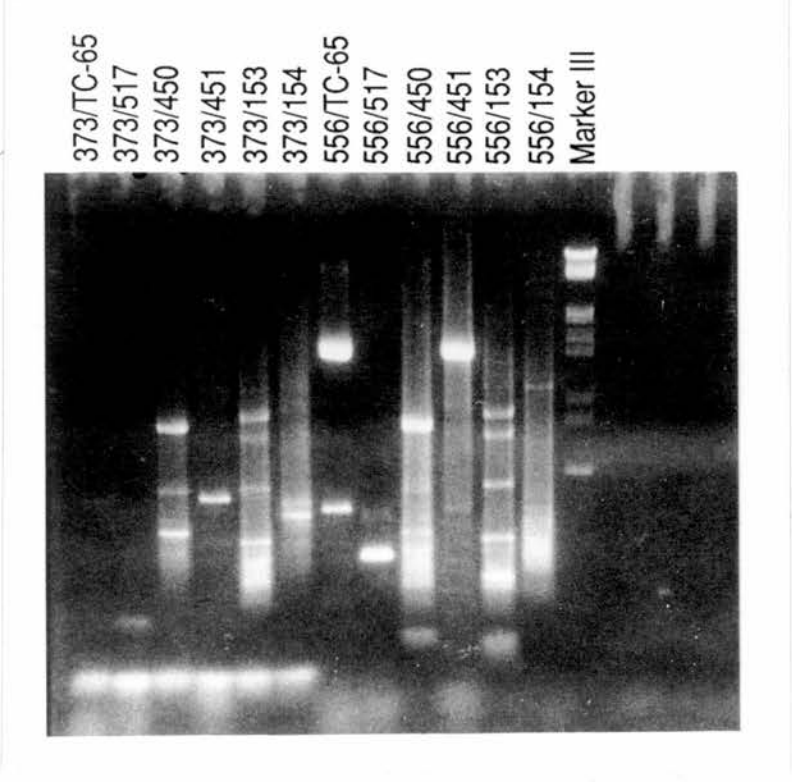
Autoradiographs of dot blots prepared with adult and foetal kidney cDNA-32A-A4 YAC hybrid fractions and hybridised with either labelled 32A-A4 total YAC DNA (Figure 5.23a) or with labelled *Alu* fragment (Figure 5.23b). All the fragments hybridised to both the YAC and the *Alu* sequence.

5.3.8 End cloning

32A-A4 (26-6/VK5-positive human YAC)

Attempts were made to clone the ends of 32A-A4 by both the asymmetric PCR and the *Alu*-vector PCR methods. 32A-A4 was digested with *Bgl*II and ligated to catch-linkers, or digested with *Rsa*I and ligated to catch-linkers. Both were amplified by PCR with 962L primers and used as starting material for the asymmetric PCR. No consistent product was obtained using this method. *Alu*-vector PCR was performed using vector primers 373 and 556 with each of the *Alu* primers in turn. Most of the pairs of primers produced multiple bands, but pairs 373/451, 373/154 and 556/517 gave a single band (Figure 5.24).

Figure 5.24



Gel photograph showing the PCR products obtained from *Alu*-vector PCR of YAC 32A-A4 with different sets of primers. Each of the six *Alu* primers (TC-65, 517, 450, 451, 153 and 154) were combined in turn with either the left arm vector primer 373 or the right arm vector primer 556. Pairs 373/451, 373/154 and 556/517 gave single bands. Pair 373/TC-65 yielded no PCR product, and all the other pairs gave multiple bands.

These three bands were re-amplified by PCR of a needle stab of DNA from the agarose gel, and pairs 373/154 and 556/517 again gave single bands. Both fragments were isolated from a 0.8% TAE agarose gel, purified using Geneclean and probed against a filter from a pulsed field gel containing 32A-A4 and control YACs. Fragment 373/154 hybridised strongly to 32A-A4 in a position consistent with the YAC band, and also hybridised weakly to the control clones, again over the site of the YAC bands (Figure 5.25). This cross hybridisation is likely to be due to the presence of a small amount of vector sequence present in the end-fragment. Fragment 556/517 gave a similar picture. Both fragments were subcloned into pBluescript ready for sequencing.

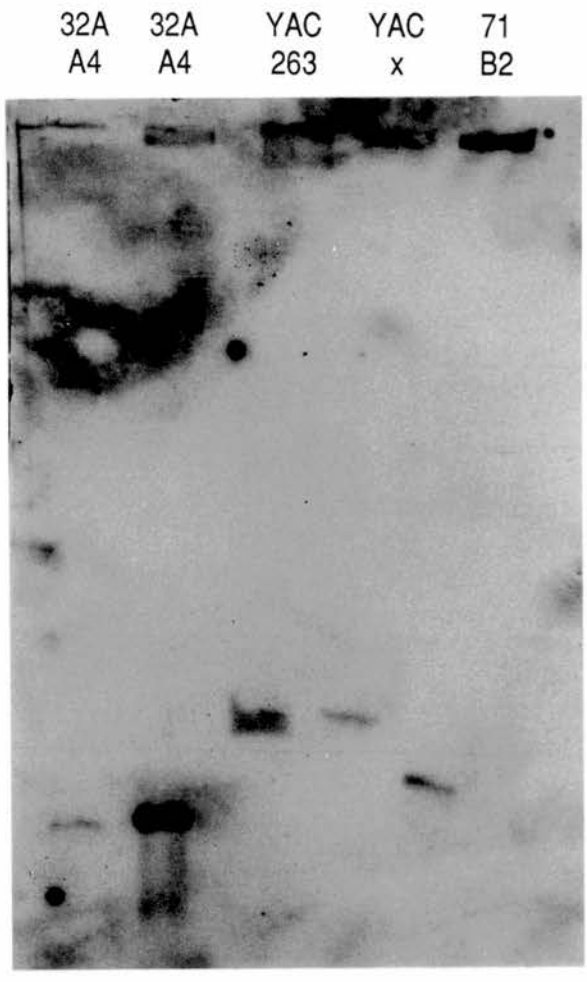
C4-I12 (ATPL-positive mouse YAC)

Attempts were made to clone the ends of C4-I12 by the asymmetric PCR method. C4-I12 digested with *Bgl*II and ligated to catch-linkers, and C4-I12 digested with *Rsa*I and ligated to catch-linkers, were both amplified by PCR with 962L primers and used as starting material for the asymmetric PCR. No consistent product was obtained using this method.

5.3.9 Screening of a mouse cosmid library

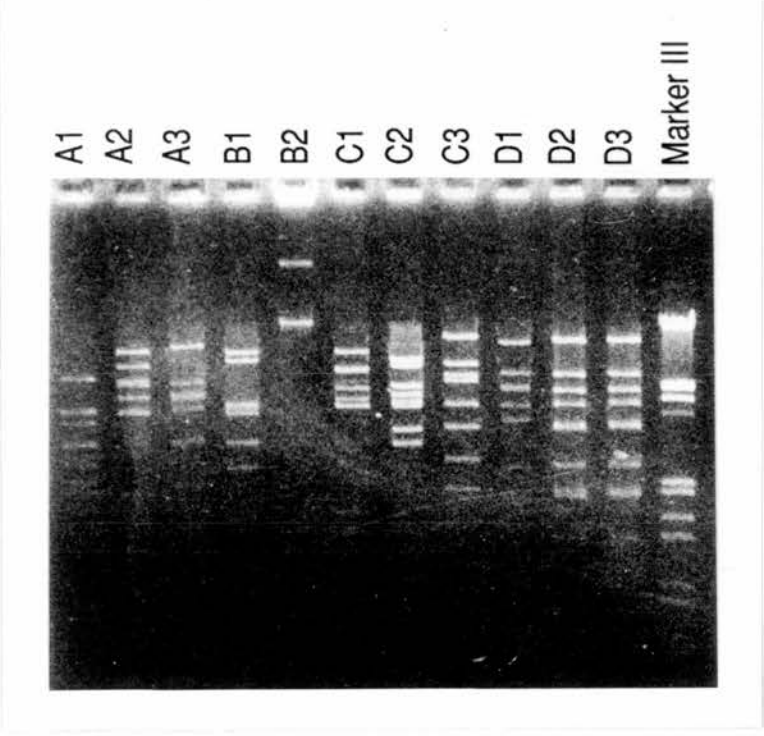
A mouse cosmid library was obtained from Dr. J. H. McVey and was screened with YAC C4-I12, using amplified catch-linkered DNA as a probe and mouse *Col*I DNA as competitor. Four positive colonies (A, B, C, and D) were identified on screening the master filters, and these were picked and re-plated. Secondary screening was performed on the 1:1000 dilutions of the above colonies using the same protocol. The filters from colonies A and C were covered with positive signals corresponding to the colonies, and plates B and D showed a few positive colonies. Positive colonies from each plate were grown up and DNA prepared using the rapid boiling miniprep technique. The DNA was digested with *Eco*RI, electrophoresed in a 0.8% agarose gel (Figure 5.26) and transferred to a filter.

Figure 5.25



Autoradiograph showing hybridisation of the 32A-A4 end clone 373/154 to various YACs. Clone 373/154 hybridises strongly to 32A-A4 in a position consistent with the YAC band, and also hybridises weakly to the control clones, again over the site of the YAC bands.

Figure 5.26



*Eco*RI digests of cosmid clones, obtained by screening a mouse cosmid library directly with mouse YAC C4-I12 which is positive for the ATPL locus. Clones D1, D2, and D3 appear to be the same and are therefore more likely to be genuine.

The digests suggest that several different cosmids have been obtained, some of which appear to be the same. Probing of the filter with labelled C4-I12 gives a strong positive signal in all lanes. The cosmids now need to be probed back to the YAC to ensure that they contain sequence from C4-I12, and they will then be a useful resource for screening cDNA libraries and for repeating attempts at localisation by fluorescence *in situ* hybridisation.

5.4 DISCUSSION

Identification of cDNAs and candidate genes requires cloned DNA from the region of search. The region around the PKD1 gene on chromosome 16, from GGG1 (D16S259) to 26-6PROX (D16S125), has been almost entirely cloned in a cosmid contig, although two 'unclonable' gaps of <20 kb and 60 kb remain (Germino *et al.*, 1992). Since YACs are able to accommodate up to 1 Megabase of insert DNA, it is potentially much quicker to build up a YAC contig of an area than a cosmid contig. In addition, segments of the human genome that have been cloned into YACs are generally stable, and nearly all DNA appears to be clonable (Abidi *et al.*, 1990). It was therefore decided to make use of YAC technology in order to obtain cloned DNA from the PKD1 region. As discussed in Chapter 4, the finding of linkage disequilibrium between PKD1 and VK5 (D16S94), argued that the search for PKD1 should be concentrated in the proximal half of the 750 kb region between GGG1 (D16S259) and 26-6 (D16S125), and extended proximally towards VK5 (D16S94). Primers to the GGG1 (D16S259), 26-6 (D16S125) and SM33p22 loci were already available as part of a European Concerted Action programme (Dr. M. H. Breuning). In view of the disequilibrium findings, primers were designed to the VK5 (D16S94) and ATPL loci, so that four proximal loci and one distal locus were used to screen the YAC libraries.

5.4.1 GC content of VK5

In order to design primers to the VK5 (D16S94) locus, the VK5 insert (in pBR322) was subcloned and sequenced. The DNA surrounding the α-globin complex is known to be GC-rich (Fischel-Ghodsian *et al.*, 1987), and Harris *et al.* (1990) found that this GC-rich region extended for more than 2 Megabases proximally. A large number of CpG clusters have been mapped to the interval GGG1 (D16S259) - 26-6 (D16S125) suggesting that the PKD1 region will be gene-rich (Germino *et al.*, 1992). Mammalian genomes have an average G + C content of 41%. In subcloning and sequencing VK5, five small fragments of approximately 200 bp were sequenced. One of the ends of the complete insert, adjacent to the pBR322 *Eco*RI site, has a slightly

higher than average GC content at 49%, but the other end of the insert (adjacent to the pBR322 *Bam*HI site) and the ends of the subclones have a very high GC content of between 62 and 64%. This therefore indicates that the GC-rich region, and probably the gene-rich region, extends beyond 26-6 (D16S125) at least to VK5 (D16S94).

5.4.2 Screening of YAC libraries

There are two basic methods of screening YAC libraries for the presence of single copy sequences. Hybridisation with radioactively labelled probes (Brownstein *et al.*, 1989) requires that the individual YAC clones have been transferred to filters. Screening by this method can yield false positives by cross-hybridisation, especially since probes may contain repeat sequences found in a number of YACs. As an alternative, a PCR-based method has been developed (Green and Olson, 1990) in which initial screening was by PCR of pools of YAC DNA, and the individual YAC clone(s) identified by a final hybridisation step. This method of screening requires that every probe be sequenced and primer pairs synthesised. The generation of inappropriate PCR products has been observed with several PCR assays, presumably generated from sequences that match the PCR primers imperfectly (Green and Olson, 1990). Green and Olson (1990) compared screening by colony hybridisation with that by PCR, and detected the same positive clones for 8 different genes. Screening by colony hybridisation, however, also yielded numerous false positives.

Since both the CEPH and ICI YAC libraries were available as pools of YAC DNA, and since some PCR primers were already available, it was decided to use PCR screening as the method of choice in this project. Rows x columns are now also available, enabling the final round of screening to be performed by PCR also.

In a library of approximately 23,000 clones with an average insert size of 250 kb, single-copy sequences are expected, on average, to be represented twice (Green and Olson, 1990). Similarly, a complexity of 3-4 genome equivalents should give a 95% chance of finding a given sequence, increasing to 98-99% with 6-10 genome

equivalents (Riley *et al.*, 1992). The ICI YAC library consists of 35,000 clones with an average insert size of 350 kb, and the CEPH YAC library has 50,000 clones with a mean insert size of 430 kb. The complexity of these two libraries is 3.5 and 7 genome equivalents respectively. Screening of these libraries with a single-copy probe would therefore be expected to yield several positive YAC clones in 95-98% of cases. In this project, the CEPH YAC library was screened with primers to three loci [GGG1 (D16S259), ATPL and VK5 D16S94)], and only one YAC clone was obtained. The ICI YAC library was screened with primers to five loci [GGG1 (D16S259), ATPL, SM33p22, 26-6 (D16S125) and VK5 D16S94)] and only two YAC clones were obtained. This is a far lower return than would be expected from the above data. In order to ensure that positive clones had not been missed, the sensitivity of screening was increased by Southern blotting the PCR reaction products and by probing the filters with labelled PCR product from human genomic DNA. Green and Olson (1990) noted complete inhibition of PCR in the presence of more than 60-80 ng of purified yeast DNA/µl, the optimal concentration for PCR being 20 ng/µl. Care was therefore taken not to exceed this concentration.

M. H. Breuning and P. Harris (personal communication) have since also reported little success in isolating YACs using probes from the same region. It therefore appears that 16p13.3 is under-represented in these two libraries, perhaps because YAC clones from this region are inherently unstable. It is of interest that one of the three YACs isolated was potentially chimaeric, and another was apparently unstable on growing up in culture, or contained sequences harmful to the host.

5.4.3 Co-cloning

Chimaeric YACs contain insert fragments from non-contiguous regions of the genome. This co-cloning may potentially occur as a result of co-ligation of two fragments of insert DNA, or where co-transformation of more than one ligated DNA recombinant in the same cell is followed by homologous recombination between repetitive sequences (such as *Alu* sequences) in the two molecules (Green *et al.*,

1991). In order to detect the presence of chimaerism in a YAC, Green *et al.* (1991) developed sequence tagged sites (STSs) from the ends of the YAC inserts, and the assigned these STSs to particular human chromosomes by PCR analysis of DNA from a panel of human-hamster hybrid cell lines. This was then confirmed using λ subclones containing the ends of the YAC inserts as probes in fluorescence *in situ* hybridisation (FISH) assays of human metaphase chromosomes. In this project, the whole YAC was labelled by digestion, attachment of catch-linkers, amplification and biotinylation, and this material was used for FISH analysis of human metaphase chromosomes. Co-cloning has been measured in the range of 50% for contigs in the CF (Green and Olson, 1990) and HLA (Bronson *et al.*, 1991) regions using YACs from a library of total human DNA. The assembly of contigs is usually unaffected providing that a sufficient number of non-chimaeric clones are available to span the gap. YAC libraries made from human-rodent hybrid cells seem to show a lower level (20%) of co-cloning (Little *et al.*, 1992). The potentially chimaeric YAC 71B2/71B4 obtained from the CEPH library in this project appears unusual in that DNA segments from six different chromosomes are present: an alternative explanation might be that it contains a low copy repetitive element.

5.4.4 Clone under-representation/instability

Different reports have emphasised both the stability and instability of human YACs. Abidi *et al.* (1990) found that the recovery of Xq sequences was consistent with random cloning, i.e. that all or nearly all sequences were equally clonable in YACs. Brownstein *et al.* (1989) evaluated the stability of YAC clones during prolonged propagation in yeast, and found that there was no detectable change through 60 generations. The human Y chromosome is rich in tandem repeats, and Neil *et al.* (1990) found that structural instability could be detected in all YAC clones examined. Albertsen *et al.* (1990) estimated the frequency of stable induced rearrangements to be no more than 5%, and Anand *et al.* (1990) mapped 15 YACs from five different loci and found one to be rearranged. Colonies from an unstable YAC usually show multiple bands that hybridise to an insert probe. After growth in

yeast, sequences contributing to the instability may be lost, and a stable deletion derivative seen. Very unstable sequences may be deleted at an early stage and are therefore essentially unclonable.

Certain YACs may also appear stable, but the structure does not correspond to that of genomic DNA. Many of the examples of instability have been found in YACs containing human repetitive DNA sequences (Neil *et al.*, 1990). The cause of most of the instability appears to be recombination carried out within the yeast cell, with the structures of rearranged YACs being those expected to result from homologous recombination events. Neil *et al.*, (1990) re-transformed YAC clones containing tandemly repeated elements into strains carrying mutations in the RAD1 and RAD52 genes. They demonstrated an increased stability of some of the unstable YAC clones when they were transformed into a host strain carrying a *rad52* mutation. The sequences most easily stabilised were those with short subunit repeats.

YACs containing subtle internal rearrangements such as inversions have also been identified (M. Dixon, personal communication). Since there is only a single yeast ARS in pYAC4, large human inserts, especially GC-rich inserts, may show delayed replication in a growing yeast culture with consequent inadvertent selection for deleted or rearranged clones which replicate faster. This could be relevant to the GC-rich region around PKD1.

It is therefore possible that chromosome 16p is a region that is inherently unstable in YACs. There have certainly been reports of repetitive sequences, with two probes known to hybridise to more than one locus on chromosome 16 (Germino *et al.*, 1992). 26-6 (D16S125) recognises two loci, 26-6PROX and 26-6DIS, less than 150 kb apart (Gillespie *et al.*, 1990), and some individuals have a third copy of the repeat sequence (P. Harris, personal communication). There is also a 40 kb segment, beginning approximately 10 kb telomeric to the *Not*I clone N54, that is present in several copies on chromosome 16 (Germino *et al.*, 1992). One of the ends of the *Kpn*I subclone of VK5 sequenced in this project also shows a repeat sequence of 28 bp. If this is the case, it may be that a YAC library transformed into a

recombination-deficient host would be more productive when screened with the same probes.

The St. Mary's Hospital Medical School mouse YAC library has been constructed in the *S cerevisiae* strain 3a, which is a *rad52* strain. This library has a complexity of 3.5 genome equivalents and therefore should yield a number of clones for each probe. There is known to be linkage conservation between human chromosome 16p and mouse chromosome 17, with many markers from between GGG1 (D16S259) and 26-6 (D16S125) mapped to this chromosome in the mouse (Himmelbauer *et al.*, 1992). This mouse YAC library was therefore screened with primers to the ATPL, VK5 (D16S94), SM33p22 and 26-6 (D16S125) loci. Despite the presence of the *rad52* host, only one YAC was successfully identified from this library.

5.4.5 Screening of cDNA libraries for transcribed sequences within YACs

Several methods have been described by which YACs can be used to screen for expressed sequences either within cDNA or YAC-derived genomic libraries. These include using the complete YAC as a hybridisation probe, enrichment by "direct selection" and exon amplification. Generally speaking, YACs are too large to be used in their entirety to screen cDNA libraries, but small YACs of 160-180 kb have been successfully used in this manner (Elvin *et al.*, 1991; Chen *et al.*, 1992). Lovett *et al.* (1991) describe a direct selection scheme based on the hybridisation of an entire cDNA library to a YAC clone immobilised on a filter. Non-specific hybrids are then eliminated and specific hybrids eluted. These molecules are then amplified and either cloned or taken through further selection/amplification cycles. Exon amplification (Buckler *et al.*, 1991) is a method for isolating exon sequences by virtue of selection for functional 5' and 3' splice sites. Fragments of cloned genomic DNA are inserted into an intron present within a mammalian expression vector and, after transfection of COS cells, cytoplasmic mRNA is screened by PCR amplification for the acquisition of an exon from the genomic fragment. The amplified exon is derived from the pairing of unrelated vector and genomic splice signals.

Another method for the selection of transcribed sequences has been described by Korn *et al.* (1992), based on the hybridisation of cDNA inserts, which had been amplified by PCR from cDNA libraries, to biotinylated DNA from cosmids. Non-specific hybrids were then removed and the selected cDNAs eluted and re-amplified by PCR. It was felt that this method could easily be modified so that catch-linkered YAC material was used rather than cosmids for screening the cDNA libraries. Korn *et al.* (1992) sheared the cosmid DNA to an average size of 500-800 bp and biotinylated this material. PCR amplification of YAC DNA digested with *Bgl*II and ligated to catch-linkers gives fragments of 300-600 bp in size, and biotinylation was achieved by the use of biotinylated primers in the PCR amplification. In a similar method to protocol B (Korn *et al.*, 1992), 600 ng biotinylated YAC DNA was pre-competed with 20 µg human *Cot*I DNA for 2½ hours. 10 µg of amplified cDNA was also pre-competed with 20 µg human *Cot*I DNA for 2½ hours, and the two were hybridised together for 20 hours. The streptavidin-coated magnetic beads were added at this stage, rather than to the biotinylated YAC. cDNAs were eluted at increasing temperatures of 75°C, 85°C and 95°C in order to try and increase specificity. Two rounds of enrichment were performed and several YAC-cDNA hybrids were obtained. Assessment of these fractions indicates that they contain a high proportion of *Alu* repetitive sequences, although at least the one fraction so far tested (AKA7) does hybridise with one band in a *Hind*III digest of the YAC that is not recognised by the *Alu* sequence. This, however, could still represent a diverged or low copy repetitive element on the cDNA. Only subcloning and sequencing of the fraction will demonstrate how much of it is unique, and how much repetitive, sequence. Given that several PCR products were obtained using *Alu* primers on 32A-A4, it seems likely that the insert DNA is *Alu*-rich. In order to try and compete out these sequences more effectively, the hybridisation could be repeated using *Alu* repetitive sequences as well as *Cot*I DNA in the pre-competition step. The enriched library could also be plated out and probed with yeast, ribosomal or mitochondrial clones, or a combination of these, and the negative clones picked for further study.

Although full restriction mapping using partial digests of the 32A-A4 YAC has not been completed, preliminary mapping of total digests with rare-cutting enzymes indicates that the human DNA insert is GC-rich, and may contain at least two CpG islands. Continued use of this YAC to identify cDNA sequences is therefore likely to be of value.

5.4.6 End-cloning

Construction of a YAC contig can be achieved either by using several probes from a region to pull out overlapping YACs, or by cloning the ends of the YAC and designing primers for further screening. It had been hoped that several overlapping clones would be identified using the primers available, but this did not prove to be the case. A number of methods have been developed to isolate YAC ends. The amp^r gene and bacterial origin of replication in the left arm of pYAC4 can be used to rescue an insert fragment attached to it, following digestion and circularisation of the total YAC DNA and transformation of the products into *E. coli*. The right arm of pYAC4 does not have the bacterial sequences necessary for plasmid rescue, but they can be inserted by retrofitting the appropriate bacterial sequences prior to plasmid rescue. Several alternatives to this cloning approach have been described. These include inverse PCR (Arveiler and Porteous, 1991), vectorette PCR (Riley *et al.*, 1991), *Alu* PCR which has been adapted for use in end-recovery by including *Alu* primers and a vector specific primer (Nelson *et al.*, 1991) and a combination of an oligo-cassette with a biotinylated primer from a known locus (e.g. the vector) in conjunction with the biotin-streptavidin separation system (Rosenthal *et al.*, 1991).

Since the 32A-A4 YAC had already been digested and ligated to catch-linkers, it was decided to attempt both *Alu*-vector PCR and the biotin-streptavidin methods in order to isolate end clones. No consistent product could be obtained using the modified method of Rosenthal *et al.* (1991), most probably because the catch-linker was less specific than the oligo-cassette, and because no nested primers were available. End fragments were, however, obtained using the *Alu*-vector

PCR method, and these are available for subcloning, sequencing, and designing of PCR primers.

Since the YAC libraries were screened with primers to the SM33p22 and ATPL loci, both of which are located close to the distal end of YAC 32A-A4, and no clones were identified, it seems unlikely that screening the libraries with primers to the end clones would yield an overlapping distal clone.

5.4.7 YAC C4-I12

Having obtained a mouse YAC positive for the ATPL primers, it was of particular interest to obtain a chromosomal localisation for this YAC. After many attempts at fluorescence *in situ* hybridisation using biotinylated YAC material, a consistent signal was obtained but the mouse chromosome could not be identified. A mouse cosmid library was also screened with labelled C4-I12, and several cosmids were identified. If these are confirmed as being present on the YAC, repeat attempts at *in situ* hybridisation can be made using this cosmid material. The cosmids could also be used to screen a mouse cDNA library in order to try and identify the mouse homologue of ATPL, and any other genes present in the area.

5.5 CONCLUSION

Three YAC libraries were screened with primers to several loci surrounding PKD1, and only two stable, non-chimaeric YACs (one human and one mouse) could be identified. The human YAC has been confirmed as being located on chromosome 16p, and attempts have been made to isolate transcribed sequences present on the YAC.

CHAPTER SIX

DISCUSSION

ADPKD is one of the most common genetic diseases in man with an incidence of 1 in 1000. End-stage renal disease requiring dialysis or transplantation occurs in at least 50% of patients by the age of 70 years (Churchill *et al.*, 1984; Parfrey *et al.*, 1990), despite treatment for hypertension. Isolation of the genes responsible for ADPKD would make a significant contribution to the understanding of the pathophysiology of the disease, would allow accurate diagnosis, and would point to possible treatment options.

Linkage was first demonstrated between ADPKD and 3'HVR (D16S85) in 1985 (Reeders *et al.*, 1985) but, despite intensive effort since then, the PKD1 gene has yet to be cloned. This is likely to be due to several factors. Firstly, the biochemical defect is unknown and therefore a positional cloning approach is required. Secondly, there are no known cytogenetic abnormalities in PKD1 patients to aid chromosomal localisation of the gene. Thirdly, genetic heterogeneity has been shown to be present, and care therefore has to be taken in order to distinguish families linked to chromosome 16 (PKD1) from those which are unlinked (PKD2). Recently, PKD2 has been shown to be linked to chromosome 4 markers (Peters *et al.*, 1993). Fourthly, genetic mapping techniques, such as linkage analysis, identification of recombinants and assessment of linkage disequilibrium, all provide guides to the chromosomal location, but can be misleading, as was the case in Huntington's Disease. Finally, the area is very GC-rich (Harris *et al.*, 1990) and therefore likely to contain several genes (Germino *et al.*, 1992), all of which require to be screened for mutations.

The aim of this project was initially to establish the presence or absence of genetic heterogeneity, and to verify the chromosomal location of PKD1 in the Scottish population by means of linkage analysis, recombinant mapping, and linkage disequilibrium mapping. Once this was achieved, it was anticipated that

yeast artificial chromosome (YAC) technology would allow rapid identification of cloned DNA from the region which could then be used to screen cDNA libraries and identify candidate genes.

The presence of genetic heterogeneity was confirmed in the Scottish population with the maximum likelihood value of α being 0.81 (81% of families linked to chromosome 16p), and 3 out of 35 families (8.6%) identified as being unlinked after haplotype analysis. The presence of genetic heterogeneity is important since the accurate localisation of PKD1 requires a genetically homogeneous population, and it is important that mutation screening is performed using individuals from families that are clearly linked to 16p.

Multipoint linkage analysis of our set of families identified the most likely location of PKD1 as:-

D16S84-0.004-PKD1-0.006-D16S283-0.005-D16S125-0.005-D16S94-cen

This localisation is entirely in keeping with previous multipoint studies (Reeders *et al.*, 1988; Breuning *et al.*, 1990a; Germino *et al.*, 1990). Identification of recombinants, one between PKD1 and CMM65 (D16S84) and three between PKD1 and 26-6 (D16S125), localised the position of PKD1 to between CMM65 (D16S84) and 26-6 (D16S125). This is again consistent with the majority of previous recombinant data (Somlo *et al.*, 1992b), although it should be noted that one recombinant in a 33 year old unaffected individual is at variance with the rest of the data. Conflicting localisations by recombinant mapping also occurred in Huntington's Disease (MacDonald *et al.*, 1989). In this case, attempts were made to sort out the anomaly using linkage disequilibrium mapping. The picture of linkage disequilibrium seen in Huntington's Disease was extremely complicated, but a common segment could be identified by haplotype analysis (MacDonald *et al.*, 1992). This segment was later shown to contain the gene (The Huntington's Disease Collaborative Research Group, 1993) and linkage disequilibrium mapping therefore proved useful in this situation.

In this study, linkage disequilibrium was first detected between PKD1 and VK5 (D16S94). VK5 (D16S94) is located on the proximal side of the CMM65

(D16S84)-26-6 (D16S125) segment thought to contain the gene by the majority of recombinant mapping data. Further markers were therefore studied, and linkage disequilibrium was detected between PKD1 and CW4/cKLH9 (D16S291), and between PKD1 and W5.2. This would be compatible with a plateau of linkage disequilibrium covering the proximal half of the CMM65 (D16S84)-26-6 (D16S125) interval and extending at least to VK5 (D16S94). Such a plateau of linkage disequilibrium, within which not all markers demonstrate disequilibrium, has also been seen in cystic fibrosis (Kerem *et al.*, 1989). This finding would argue that the search for PKD1 should be concentrated in the proximal half of the CMM65 (D16S84)-26-6 (D16S125) region, and perhaps should be extended as far as VK5 (D16S94) or beyond. Another possible explanation for the above findings is that PKD1 may be a large gene spanning both 26-6 (D16S125) and VK5 (D16S94), the 26-6 (D16S125) recombinants being intragenic.

It is of interest that a very promising candidate gene, AJ1, has been excluded on the grounds of recombinant mapping data and the absence of a mutation in one affected individual (Gillespie *et al.*, 1991). It is assumed that the screened DNA is from an individual with a PKD1 mutation, although no data is provided on the linkage analysis of this family. Should the screened individual be from an unlinked family, however, then AJ1 remains a strong candidate gene despite the recombinant mapping data, particularly in view of the linkage disequilibrium data presented here.

Further genetic mapping is unlikely to narrow the localisation of PKD1 further, and attempts were therefore made to obtain cloned DNA from 16p13.3 by screening of yeast artificial chromosome (YAC) libraries. These attempts were largely unsuccessful with only four YACs being isolated from three YAC libraries (two human and one mouse). Only two of these four YACs were able to be used, since one was unstable and one either highly chimaeric or containing low-copy repetitive sequences. The human YAC 32A-A4, positive for both 26-6 (D16S125) and VK5 (D16S94), was from the area localised by genetic mapping as being likely to contain the PKD1 gene, and it also contained at least two CpG clusters suggesting that

genes were present on the YAC. Initial attempts to isolate cDNA from a human and a foetal kidney cDNA library using a YAC-cDNA hybridisation technique were unsuccessful. 16p13.3 is thought to lie in a T band which is both *Alu*-rich and GC-rich (Holmquist, 1992). Certainly, the sequence data obtained in this project confirms that the area around VK5 is very GC rich and one problem with the YAC-cDNA hybridisation technique was the presence of a high number of *Alu* repeat sequences. The technique could be modified for future use by increasing the competition, especially for *Alu* repeat sequences. It would also be worthwhile using the whole YAC as a probe against the cDNA libraries

The region of 16p surrounding PKD1 is very GC-rich and likely to contain a large number of genes. Several cDNAs have already been isolated (Somlo *et al.*, 1992a) and screening for mutations therefore presents a formidable task. Once a candidate gene has been isolated, the knowledge of its function will prove useful in solving the mystery surrounding cyst formation. Many experiments on the pathophysiology of the disease have been performed, and it seems clear that a combination of abnormal cell proliferation and abnormal fluid transport is required for cyst formation. Epithelial proliferation can be the result of alteration of extracellular matrix composition, altered susceptibility to, or synthesis of, peptide growth factors, or abnormal expression of growth-related genes including proto-oncogenes. Abnormal fluid transport could be the result of alterations in ion pumping, or reflect changes in polarity of cell structure or function. Since some of the structural proteins involved in the extracellular matrix, e.g. E-cadherin and N-cell adhesion molecule, seem to be involved in the polarisation of renal epithelial cells, it is possible that a defect in a structural protein may account for both epithelial cell proliferation and abnormal ion and fluid transport. Such a defect could also account for the extrarenal manifestations of polycystic kidney disease.

One of the striking features of ADPKD is the absence of a detectable abnormality in most of the nephrons. Reeders (1992) has postulated that the explanation for this may involve either a two-hit model, as in retinoblastoma, or a heritable unstable repeat with somatic mosaicism of the trinucleotide repeat

number. There is, as yet, no evidence for either of these hypotheses: undoubtedly the answer will only become clear once the gene has been isolated.

In conclusion, therefore, this project has provided evidence of genetic heterogeneity of ADPKD in the Scottish population, and suggested a localisation for PKD1 on linkage analysis and recombinant mapping which is consistent with other published data. The finding of linkage disequilibrium has not been previously noted, and would argue for concentration of the search for PKD1 more proximally. A human YAC containing both the 26-6 (D16S125) and VK5 (D16S94) loci was isolated, and attempts were made to identify candidate genes using a YAC-cDNA hybridisation technique.

REFERENCES

ABIDI FE, WADA M, LITTLE RD, and SCHLESSINGER D. (1990). Yeast artificial chromosomes containing human Xq24-q28 DNA: library construction and representation of probe sequences. *Genomics* **7:** 363-376.

ALBERTSEN HM, ABDERRAHIM H, CANN HM, DAUSSET J, LE PASLIER D, and COHEN D. (1990). Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc. Natl. Acad. Sci. USA* **87:** 4256-4260.

ANAND R, RILEY JH, BUTLER R, SMITH JC, and MARKHAM AF. (1990). A 3.5 genome equivalent multi access YAC library: construction, characterisation, screening and storage. *Nucleic Acids Res.* **18:** 1951-1956.

ARVEILER B, and PORTEOUS DJ. (1991). Amplification of end fragments of YAC recombinants by inverse-polymerase chain reaction. *Technique* **3:** 24-28.

AVNER ED, SWEENEY WE Jr., and NELSON WJ. (1992). Abnormal sodium pump distribution during renal tubulogenesis in congenital murine polycystic kidney disease. *Proc. Natl. Acad. Sci. USA* **89:** 7447-7451.

BACHNER L, VINET MC, LACAVE R, BABRON MC, RONDEAU E, SRAER JD, CHEVET D, and KAPLAN J-C. (1990). Linkage study of a large family with autosomal dominant polycystic kidney disease with reduced expression. *Hum. Genet.* **85:** 221-227.

BAERT L. (1978). Hereditary polycystic kidney disease (adult form): A microdissection study of two cases at an early stage of the disease. *Kidney Int.* **13:** 519-525.

BARTON P, MALCOLM S, MURPHY C, and FERGUSON-SMITH MA. (1982). Localisation of the human α-globin gene cluster to the short arm of chromosome 16 (16p12-16pter) by hybridisation in situ. *J. Mol. Biol.* **156:** 269-278.

BEAR JC, PARFREY PS, MORGAN J, CRAMER BC, McMANAMON PJ, GAULT MH, CHURCHILL DN, SINGH M, HEWITT R, SOMLO S, and REEDERS ST. (1989). Autosomal dominant polycystic kidney disease: ultrasonographic detection and prognosis of PKD1 and PKD2 forms. [Abstract] *Am. J. Hum. Genet.* **45:** A39.

BEAR JC, PARFREY PS, MORGAN JM, MARTIN CJ, and BENVON C. (1992). Autosomal dominant polycystic kidney disease: new information for genetic counselling. *Am. J. Med. Genet.* **43:** 548-553.

BELL PE, HOSSACK KF, GABOW PA, DURR JA, JOHNSON AM, and SCHRIER RW. (1988). Hypertension in autosomal dominant polycystic kidney disease. *Kidney Int.* **34:** 683-690.

BILLER DS, and PFLUEGER SMV. (1990). A feline model for autosomal dominant adult polycystic kidney disease. [Abstract] *Am. J. Hum. Genet.* **47:** A48.

BIRENBOIM N, DONOSO VS, HUSEMAN RA and GRANTHAM JJ. (1987). Renal excretion and cyst accumulation of β_2 microglobulin in polycystic kidney disease. *Kidney Int.* **31:** 85-92.

BOULTER CA, AGUZZI A, EVANS MJ, and AFFARA N. (1992). A chimaeric mouse model for autosomal dominant polycystic kidney disease. In Breuning MH, Devoto M, and Romeo G (eds): *Polycystic kidney disease. Contrib. Nephrol.* Basel, Karger **97:** 60-70.

BREEN M, ARVEILER B, MURRAY I, GOSDEN JR, and PORTEOUS DJ. (1992). YAC mapping by FISH using *Alu*-PCR-generated probes. *Genomics* **13:** 726-730.

BREUNING MH, MADAN K, VERJAAL M, WIJNEN JT, MEERA KHAN P, and PEARSON PL. (1987a). Human α-globin maps to pter-p13.3 in chromosome 16 distal to PGP. *Hum. Genet.* **76:** 287-289.

BREUNING MH, REEDERS ST, BRUNNER H, IJDO JW, SARIS JJ, VERWEST A, VAN OMMEN GJB, and PEARSON PL. (1987b). Improved early diagnosis of adult polycystic kidney disease with flanking DNA markers. *Lancet* **ii:** 1359-1361.

BREUNING MH, SNIJDEWINT FGM, BRUNNER H, VERWEST A, IJDO JW, SARIS JJ, DAUWERSE JG, BLONDEN L, KEITH T, CALLEN DF, HYLAND VJ, XIAO GH, SCHERER G, HIGGS DR, HARRIS P, BACHNER L, REEDERS ST, GERMINO G, PEARSON PL, and VAN OMMEN GJB. (1990a). Map of 16 polymorphic loci on the short arm of chromosome 16 close to the polycystic kidney disease gene (PKD1). *J. Med. Genet.* **27:** 603-613.

BREUNING MH, SNIJDEWINT FGM, DAUWERSE JG, SARIS JJ, BAKKER E, PEARSON PL, and VAN OMMEN GJB. (1990b). Two step procedure for early diagnosis of polycystic kidney disease with polymorphic DNA markers on both sides of the gene. *J. Med. Genet.* **27:** 614-617.

BREUNING MH, SNIJDEWINT FGM, SMITS JR, DAUWERSE JG, SARIS JJ, VAN OMMEN GJB. (1990c). A Taq I polymorphism identified by 26-6 (D16S125) proximal to the locus affecting adult polycystic kidney disease (PKD1) on chromosome 16. *Nucleic Acids Res.* **18:** 3106.

BRISSENDEN JE, ROSCOE JM, and SILVERMAN M. (1989). Linkage heterogeneity in autosomal dominant polycystic kidney disease. *Am. J. Hum. Genet.* **45:** A132.

BRONSON SK, PEI J, TAILLON-MILLER P, CHORNEY MJ, GERAGHTY DE, and CHAPLIN DD. (1991). Isolation and characterisation of yeast artificial chromosome clones linking the HLA-B and HLA-C loci. *Proc. Natl. Acad. Sci. USA* **88:** 1676-1680.

BROWNSTEIN BH, SILVERMAN GA, LITTLE RD, BURKE DT, KORSMEYER SJ, SCHLESSINGER D, and OLSON MV. (1989). Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science* **244:** 1348-1351.

BUCKLER AJ, CHANG DD, GRAW SL, BROOK JD, HABER DA, SHARP PA, and HOUSMAN DE. (1991). Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA* **88:** 4005-4009.

BURKE DT, CARLE GF, and OLSON MV. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236:** 806-812.

CALLEN DF. (1986). A mouse-human hybrid cell panel for mapping human chromosome 16. *Ann. Génét.* **29:** 235-239.

CALLEN DF, HYLAND VJ, BAKER EG, FRATINI A, GEDEON AK, MULLEY JC, FERNANDEZ KEW, BREUNING MH, and SUTHERLAND GR. (1989). Mapping the short arm of human chromosome 16. *Genomics* **4:** 348-354.

CAROTHERS AD, and WRIGHT AF. (1992). The effect of mutation on linkage disequilibrium. *Ann. Hum. Genet.* **56:** 155-158.

CHAKRAVARTI A, BUETOW KH, ANTONARAKIS SE, WABER PG, BOEHM CD, and KAZAZIAN HH. (1984). Nonuniform recombination within the human β-globin gene cluster. *Am. J. Hum. Genet.* **36:** 1239-1258.

CHARTIER FL, KEER JT, SUTCLIFFE MJ, HENRIQUES DA, MILEHAM P, and BROWN SDM. (1992). Construction of a mouse yeast artificial chromosome library in a recombination-deficient strain of yeast. *Nature Genetics* **1:** 132-136.

CHAUVEAU D, SIRIEIX M-E, SCHILLINGER F, LEGENDRE C, and GRUNFELD J-P. (1990). Recurrent rupture of intracranial aneurysms in autosomal dominant polycystic kidney disease. *Br. Med. J.* **301:** 966-967.

CHEN Z-Y, HENDRIKS RW, JOBLING MA, POWELL JF, BREAKEFIELD XO, SIMS KB, and CRAIG IW. (1992). Isolation and characterisation of a candidate gene for Norrie disease. *Nature Genetics* **1:** 204-208.

CHURCHILL DN, BEAR JC, MORGAN J, PAYNE RH, McMANAMON PJ, and GAULT MH. (1984). Prognosis of adult polycystic kidney disease re-evaluated. *Kidney Int.* **26:** 190-193.

COTO E, AGUADO S, ALVAREZ J, MENENDEZ DIAZ MJ, and LOPEZ-LARREA C. (1992). Genetic and clinical studies in autosomal dominant polycystic kidney disease type 1 (ADPKD1). *J. Med. Genet.* **29:** 243-246.

COULSON A, WATERSTON R, KIFF J, SULSTON J, and YOHARA Y. (1988). Genome linking with yeast artificial chromosomes. *Nature* **335:** 184-186.

COWLEY BD Jr, GUDAPATY S, KRAYBILL AL, BARASH BD, HARDING MA, CALVET JP, and GATTONE II, VH. (1993). Autosomal-dominant polycystic kidney disease in the rat. *Kidney Int.* **43:** 522-534.

COWLEY BD Jr., SMARDO FL Jr., GRANTHAM JJ, and CALVET JP. (1987). Elevated c-*myc* protooncogene expression in autosomal recessive polycystic kidney disease. *Proc. Natl. Acad. Sci. USA* **84:** 8394-8398.

CRAIG JM, and BICKMORE WA. (1993). Chromosome bands-flavours to savour. *BioEssays* **15:** 349-354.

DALGAARD OZ. (1957). Bilateral polycystic disease of the kidneys. A follow-up of two hundred and eighty-four patients and their families. *Acta Med. Scand.* **158 (Suppl. 328):** 1-233.

DAVISSON MT, GUAY-WOODFORD LM, HARRIS HW, and D'EUSTACHIO P. (1991). The mouse polycystic kidney disease mutation (*cpk*) is located on proximal chromosome 12. *Genomics* **9:** 778-781.

DONIS-KELLER H, GREEN P, HELMS C, CARTINHOUR S, WEIFFENBACH B, STEPHENS K, KEITH TP, *ET AL.* (1987). A genetic linkage map of the human genome. *Cell* **51:** 319-337.

DRYJA TP, MCGEE TL, REICHEL E, HAHN LB, COWLEY GS, YANDELL DW, SANDBERG MA, and BERSON EL. (1990). A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. *Nature* **343:** 364-366.

DUNNILL MS, MILLARD PR, and OLIVER D. (1977). Acquired cystic disease of the kidneys: a hazard of long-term intermittent maintenance haemodialysis. *J. Clin. Path.* **30:** 868-877.

ELLES RG. (1992). Linkage disequilibrium between D16S94 and the locus for adult polycystic kidney disease. [Letter] *J. Med. Genet.* **29:** 758.

ELLES RG, READ AP, HODGKINSON KA, WATTERS A and HARRIS R. (1990). Recombination or heterogeneity: is there a second locus for adult polycystic kidney disease? *J. Med. Genet.* **27:** 413-417.

ELVIN P, SLYNN G, BLACK D, GRAHAM A, BUTLER R, RILEY J, ANAND R, and MARKHAM AF. (1990). Isolation of cDNA clones using yeast artificial chromosome probes. *Nucleic Acids Res.* **18:** 3913-3917.

ESTIVILL X, SCAMBLER PJ, WAINWRIGHT BJ, HAWLEY K, FREDERICK P, SCHWARTZ M, BAIGET M, KERE J, WILLIAMSON R, and FARRALL M. (1987). Patterns of polymorphism and linkage disequilibrium for cystic fibrosis. *Genomics* **1:** 257-263.

FISCHEL-GHODSIAN N, NICHOLLS RD, and HIGGS DR. (1987). Long range genome structure around the human α-globin complex analysed by PFGE. *Nucleic Acids Res.* **15:** 6197-6207.

FOSSDAL R, BÖOVARSSON M, ÁSMUNDSSON P, RAGNARSSON J, PETERS D, BREUNING MH, and JENSSON Ó. (1993). Icelandic families with autosomal dominant polycystic kidney disease: families unlinked to chromosome 16p13.3 revealed by linkage analysis. *Hum. Genet.* **91:** 609-613.

GABOW PA. (1991). Polycystic kidney disease: Clues to pathogenesis. [Editorial review] *Kidney Int.* **40:** 989-996.

GABOW PA, HEARD E, PRETORIUS D, DULEY I, BELL P, KAEHNY W, and SCHRIER R. (1987). Relationship between renal structure and hypertension in autosomal dominant polycystic kidney disease (ADPKD). [Abstract] *Kidney Int.* **31 :** 297.

GABOW PA, JOHNSON AM, KAEHNY WD, KIMBERLING WJ, LEZOTTE DC, DULEY IT, and JONES RH. (1992). Factors affecting the progression of renal disease in autosomal dominant polycystic kidney disease. *Kidney Int.* **41:** 1311-1319.

GARDNER KD. (1969). Composition of fluid in twelve cysts of a polycystic kidney. *N. Engl. J. Med.* **281:** 985-988.

GARDNER KD Jr. (1988). Cystic kidneys. *Kidney Int.* **33:** 610-621.

GERMINO GG, BARTON NJ, LAMB J, HIGGS DR, HARRIS P, XIAO GH, SCHERER G, NAKAMURA Y, and REEDERS ST. (1990). Identification of a locus which shows no genetic recombination with the autosomal dominant polycystic kidney disease gene on chromosome 16. *Am. J. Hum. Genet.* **46:** 925-933.

GERMINO GG, WEINSTAT-SASLOW D, HIMMELBAUER H, GILLESPIE GAJ, SOMLO S, WIRTH B, BARTON N, HARRIS KL, FRISCHAUF A-M, and REEDERS ST. (1992). The gene for autosomal dominant polycystic kidney disease lies in a 750-kb CpG - rich region. *Genomics* **13:** 144-151.

GILLESPIE GAJ, GERMINO GG, SOMLO S, WEINSTAT-SASLOW D, BREUNING MH, and REEDERS ST. (1990). Cosmid walking and chromosome jumping in the region of PKD1 reveal a locus duplication and three CpG islands. *Nucleic Acids Res.* **18:** 7071-7075.

GILLESPIE GAJ, SOMLO S, GERMINO GG, WEINSTAT-SASLOW D, and REEDERS ST. (1991). CpG island in the region of an autosomal dominant polycystic kidney disease locus defines the 5' end of a gene encoding a putative proton channel. *Proc. Natl. Acad. Sci. USA* **88:** 4289-4293.

GOSDEN J, HANRATTY D, STARLING J, FANTES J, MITCHELL A, and PORTEOUS D. (1991). Oligonucleotide primed *in situ* DNA synthesis (PRINS): a method for chromosome mapping, banding and investigation of sequence organisation. *Cytogenet. Cell Genet.* **57:** 100-104.

GRANTHAM JJ, GEISER JL, and EVAN AP. (1987). Cyst formation and growth in autosomal dominant polycystic kidney disease. *Kidney Int.* **31:** 1145-1152.

GRANTHAM JJ, UCHIC M, CRAGOE EJ Jr., KORNHAUS J, GRANTHAM JA, DONOSO V, MANGOO-KARIM R, EVAN A, and McATEER J. (1989). Chemical modification of cell proliferation and fluid secretion in renal cysts. *Kidney Int.* **35:** 1379-1389.

GREEN ED, and OLSON MV. (1990). Systematic screening of yeast artificial chromosome libraries by use of the polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **87:** 1213-1217.

GREEN ED, RIETHMAN HC, DUTCHIK JE, and OLSON MV. (1991). Detection and characterisation of chimeric yeast artificial chromosome clones. *Genomics* **11:** 658-669.

GRETZ N, HOCKER A, BAUR S, LASSERRE JJ, BACHMANN S, WALDHERR R, and STRAUCH M. (1992). Rat models of polycystic kidney disease. In Breuning MH, Devoto M, and Romeo G (eds): *Polycystic kidney disease. Contrib. Nephrol.* Basel, Karger **97:** 35-46.

HARRIS PC, BARTON NJ, HIGGS DR, REEDERS ST, and WILKIE AOM. (1990). A long-range restriction map between the α-globin complex and a marker closely linked to the polycystic kidney disease 1 (PKD1) locus. *Genomics* **7:** 195-206.

HARRIS PC, THOMAS S, RATCLIFFE PJ, BREUNING MH, COTO E, and LOPEZ - LARREA C. (1991). Rapid genetic analysis of families with polycystic kidney disease 1 by means of a microsatellite marker. *Lancet* **338:** 1484-1487.

HATFIELD PM, and PFISTER RC. (1972). Adult polycystic disease of the kidneys (Potter type 3). *J. Am. Med. Assoc.* **222:** 1527-1531.

HEGELE RA, PLAETKE R, and LALOUEL J-M. (1990). Linkage disequilibrium between DNA markers at the low-density lipoprotein receptor gene. *Genet. Epidemiol.* **7:** 69-81.

HIMMELBAUER H, GERMINO GG, CECCHERINI I, ROMEO G, REEDERS ST, and FRISCHAUF A-M. (1991). Saturating the region of the polycystic kidney disease gene with *Not*I linking clones. *Am. J. Hum. Genet.* **48:** 325-334.

HIMMELBAUER H, POHLSCHMIDT M, SNAREY A, GERMINO GG, WEINSTAT-SASLOW D, SOMLO S, REEDERS ST, and FRISCHAUF A-M. (1992). Human-mouse homologies in the region of the polycystic kidney disease gene (PKD1). *Genomics* **13:** 35-38.

HOLMES DS, and QUIGLEY M. (1981). A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* **114:** 193-197.

HOLMQUIST GP. (1992). Chromosome bands, their chromatin flavours and their functional features. [Review] *Am. J. Hum. Genet.* **51:** 17-37.

HULTEN M. (1988). Linkage heterogeneity in autosomal dominant polycystic kidney disease. *Lancet* **ii:** 451-452.

HYLAND VJ, SUTHERS GK, FRIEND K, MacKINNON RN, CALLEN DF, BREUNING MH, KEITH T, BROWN VA, PHIPPS P, and SUTHERLAND GR. (1990). Probe, VK5B, is located in the same interval as the autosomal dominant polycystic kidney disease locus, PKD1. *Hum. Genet.* **84:** 286-288.

JARMAN AP, NICHOLLS RD, WEATHERALL DJ, CLEGG JB, and HIGGS DR. (1986). Molecular characterisation of a hypervariable region downstream of the human α-globin gene cluster. *EMBO J.* **5:** 1857-1863.

KAEHNY W, BELL P, EARNEST M, STEARS J, and GABOW P. (1987). Family clustering of intracranial aneurysms in autosomal dominant polycystic kidney disease. [Abstract] *Kidney Int.* **31:** 204.

KEITH TP, GREEN P, REEDERS ST, BROWN VA, PHIPPS P, BRICKER A, FALLS K, REDIKER KS, POWERS JA, HOGAN C, NELSON C, KNOWLTON R, and DONIS-KELLER H. (1990). Genetic linkage map of 46 DNA markers on human chromosome 16. *Proc. Natl. Acad. Sci. USA* **87:** 5754-5758.

KEREM B-S, ROMMENS JM, BUCHANAN JA, MARKIEWICZ D, COX TK, CHAKRAVARTI A, BUCHWALD M, and TSUI L-C. (1989). Identification of the cystic fibrosis gene: genetic analysis. *Science* **245:** 1073-1080.

KIMBERLING WJ, FAIN PR, KENYON JB, GOLDGAR D, SUJANSKY E, and GABOW PA. (1988). Linkage heterogeneity of autosomal dominant polycystic kidney disease. *N. Engl. J. Med.* **319:** 913-918.

KNUDSON AG Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. USA* **68:** 820-823.

KORN B, SEDLACEK Z, MANCA A, KIOSCHIS P, KONECKI D, LEHRACH H, and POUSTKA A. (1992). A strategy for the selection of transcribed sequences in the Xq28 region. *Hum. Molec. Genet.* **1:** 235-241.

KUMAR S, KIMBERLING WJ, GABOW PA, and KENYON JB. (1991). Genetic linkage studies of autosomal dominant polycystic kidney disease: search for the second gene in a large Sicilian family. *Hum. Genet.* **87:** 129-133.

KUNKEL LM, TANTRAVAHI U, EISENHARD M, and LATT SA. (1982). Regional localisation on the human X of DNA segments cloned from flow sorted chromosomes. *Nucleic Acids Res.* **10:** 1557-1578.

LAMB J, WILKIE AOM, HARRIS PC, BUCKLE VJ, LINDENBAUM RH, BARTON NJ, REEDERS ST, WEATHERALL DJ, and HIGGS DR. (1989). Detection of breakpoints in submicroscopic chromosomal translocation, illustrating an important mechanism for genetic disease. *Lancet* **ii:** 819-824.

LATHROP GM, LALOUEL J-M, JULIER C, and OTT J. (1984). Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci. USA* **81:** 3443-3446.

LAZAROU LP, DAVIES F, SARFARAZI M, COLES GA, and HARPER PS. (1987). Adult polycystic kidney disease and linked RFLPs at the α-globin locus: a genetic study in the South Wales population. *J. Med. Genet.* **24:** 466-473.

LEIER CV, BAKER PB, KILMAN JW, and WOOLEY CF. (1984). Cardiovascular abnormalities associated with adult polycystic kidney disease. *Ann. Intern. Med.* **100:** 683-688.

LEVEY AS, PAUKER SG, and KASSIRER JP. (1983). Occult intracranial aneurysms in polycystic kidney disease: When is cerebral arteriography indicated? *N. Engl. J. Med.* **308:** 986-994.

LEWONTIN RC. (1988). On measures of gametic disequilibrium.*Genetics* **120:** 849-852.

LINDSAY S, and BIRD AP. (1987). Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature* **327:** 336-338.

LITT M, and JORDE LB. (1986). Linkage disequilibria between pairs of loci within a highly polymorphic region of chromosome 2q. *Am. J. Hum. Genet.* **39:** 166-178.

LITTLE RD, PILIA G, JOHNSON S, D'URSO M, and SCHLESSINGER D. (1992). Yeast artificial chromosomes spanning 8 megabases and 10-15 centimorgans of human cytogenetic band Xq26. *Proc. Natl. Acad. Sci. USA* **89:** 177-181.

LOVETT M, KERE J, and HINTON LM. (1991). Direct selection: A method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. USA* **88:** 9628-9632.

MACDONALD ME, HAINES JL, ZIMMER M, CHENG SV, YOUNGMAN S, WHALEY WL, WEXLER N, BUCAN M, ALLITTO BA, SMITH B, LEAVITT J, POUSTKA A, HARPER P, LEHRACH H, WASMUTH JJ, FRISCHAUF A-M, and GUSELLA JF. (1989). Recombination events suggest potential sites for the Huntungton's Disease gene. *Neuron* **3:** 183-190.

MACDONALD ME, LIN C, SRINIDHI L, BATES G, ALTHERR M, WHALEY WL, LEHRACH H, WASMUTH J, and GUSELLA JF. (1991). Complex patterns of linkage disequilibrium in the Huntington Disease region. *Am. J. Hum. Genet.* **49:** 723-734.

MACDONALD ME, NOVELLETTO A, LIN C, TAGLE D, BARNES G, BATES G, TAYLOR S, ALLITTO B, ALTHERR M, MYERS R, LEHRACH H, COLLINS FS, WASMUTH JJ, FRONTALI M, and GUSELLA JF. (1992). The Huntington's Disease candidate region exhibits many different haplotypes. *Nature Genetics* **1:** 99-103.

MACKAY K, STRIKER LJ, PINKERT CA, BRINSTER RL, and STRIKER GE. (1987). Glomerulosclerosis and renal cysts in mice transgenic for the early region of SV40. *Kidney Int.* **32:** 827-837.

MACNICOL AM, WRIGHT AF, and WATSON ML. (1991). Education and attitudes in families with adult polycystic kidney disease. *Nephrol. Dial. Transplant.* **6:** 27-30.

MANDICH P, RESTAGNO G, NOVELLI G, BELLONE E, POTENZA L, VARETTO O, DALLAPICCOLA B, CARBONARA A, and AJMAR F. (1990). Autosomal dominant polycystic kidney disease: A linkage evaluation of heterogeneity in Italy. *Am. J. Med. Genet.* **35:** 579-581.

MANGOO-KARIM R, UCHIC M, LECHENE C, and GRANTHAM JJ. (1989). Renal epithelial cyst formation and enlargement in vitro: dependence on cAMP. *Proc. Natl. Acad. Sci. USA* **86:** 6007-6011.

MILLS KA, BUETOW KH, XU Y, WEBER JL, ALTHERR MR, WASMUTH JJ, and MURRAY JC. (1992). Genetic and physical maps of human chromosome 4 based on dinucleotide repeats. *Genomics* **14:** 209-219.

MILUTINOVIC J, FIALKOW PJ, AGODOA LY, PHILLIPS LA, RUDD TG, and BRYANT JI. (1984). Autosomal dominant polycystic kidney disease: symptoms and clinical findings. *Q. J. Med.* **53:** 511-522.

MILUTINOVIC J, FIALKOW PJ, RUDD TG, AGODOA LY, PHILLIPS LA, and BRYANT JI. (1980). Liver cysts in patients with autosomal dominant polycystic kidney disease. *Am. J. Med.* **68:** 741-744.

MILUTINOVIC J, FIALKOW PJ, AGODOA LY, PHILLIPS LA, and BRYANT JI. (1983). Fertility and pregnancy complications in women with autosomal dominant polycystic kidney disease. *Obstet. Gynaecol.* **61:** 566-570.

NAKAMURA Y, MARTIN C, KRAPCHO K, O'CONNELL P, LEPPERT M, LATHROP GM, LALOUEL J-M, and WHITE R. (1988). Isolation and mapping of a polymorphic DNA sequence (pCMM65) on chromosome 16 [D16S84]. *Nucleic Acids Res.* **16:** 3122.

NEIL DL, VILLASANTE A, FISHER RB, VETRIE D, COX B, and TYLER-SMITH C. (1990). Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. *Nucleic Acids Res.* **18:** 1421-1428.

NELSON DL, LEDBETTER SA, CORBO L, VICTORIA MF, RAMÍREZ-SOLIS R, WEBSTER TD, LEDBETTER DH, and CASKEY CT. (1989). *Alu* polymerase chain reaction: A method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci. USA* **86:** 6686-6690.

NØRBY S, and SCHWARTZ M. (1990). Possible locus for polycystic kidney disease on chromosome 2. [Letter] *Lancet* **ii:** 323-324.

OTT J. (1985). *Analysis of Human Genetic Linkage.* Johns Hopkins University Press, Baltimore, USA.

PARFREY PS, BEAR JC, MORGAN J, CRAMER BC, McMANAMON PJ, GAULT MH, CHURCHILL DN, SINGH M, HEWITT R, SOMLO S, and REEDERS ST. (1990). The diagnosis and prognosis of autosomal dominant polycystic kidney disease. *N. Engl. J. Med.* **323:** 1085-1090.

PETERS DJM, and SANDKUIJL LA. (1992). Genetic heterogeneity of polycystic kidney disease in Europe. In Breuning MH, Devoto M, and Romeo G (eds): *Polycystic kidney disease. Contrib. Nephrol.* Basel, Karger **97:** 128-139.

PETERS DJM, SPRUIT L, SARIS JJ, RAVINE D, SANDKUIJL LA, FOSSDAL R, BOERSMA J, VAN EIJK R, NØRBY S, CONSTANTINOU-DELTAS CD, PIERIDES A, BRISSENDEN JE, FRANTS RR, VAN OMMEN GJB, and BREUNING MH. (1993). Chromosome 4 localisation of a second gene for autosomal dominant polycystic kidney disease on chromosome 4. *Nature Genetics* **5:** 359-362.

PIEKE SA, KIMBERLING WJ, KENYON JB, and GABOW PA. (1989). Genetic heterogeneity of polycystic kidney disease: an estimate of the proportion of families unlinked to chromosome 16. [Abstract] *Am. J. Hum. Genet.* **45:** A58.

PIGNATELLI PM, POUND SE, CAROTHERS AD, MACNICOL AM, ALLAN PL, WATSON ML, and WRIGHT AF. (1992). Multipoint mapping of adult onset polycystic kidney disease (PKD1) on chromosome 16. *J. Med. Genet.* **29:** 638-641.

PINKEL D, STRAUME T, and GRAY JW. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridisation. *Proc. Natl. Acad. Sci. USA* **83:** 2934-2938.

POUSTKA A, RACKWITZ H-R, FRISCHAUF A-M, HOHN B, and LEHRACH H. (1984). Selective isolation of cosmid clones by homologous recombination in *Escherichia coli. Proc. Natl. Acad. Sci. USA* **81:** 4129-4133.

RAHILLY MA, SAMUEL K, ANSELL JD, MICKLEM HS, and FLEMING S. (1992). Polycystic kidney disease in the CBA/N immunodeficient mouse. *J. Pathol.* **168:** 335-342.

RAVINE D, WALKER RG, GIBSON RN, FORREST SM, RICHARDS RI, FRIEND K, SHEFFIELD LJ, KINCAID-SMITH P, and DANKS DM. (1992). Phenotype and genotype heterogeneity in autosomal dominant polycystic kidney disease. *Lancet* **ii:** 1330-1333.

REEDERS ST. (1992). Multilocus polycystic disease. *Nature Genetics* **1:** 235-237.

REEDERS ST, BREUNING MH, CORNEY G, JEREMIAH SJ, MEERA KHAN P, DAVIES KE, HOPKINSON DA, PEARSON PL, and WEATHERALL DJ. (1986). Two genetic markers closely linked to adult polycystic kidney disease on chromosome 16. *Br. Med. J.* **292:** 851-853.

REEDERS ST, BREUNING MH, DAVIES KE, NICHOLLS RD, JARMAN AP, HIGGS DR, PEARSON PL, and WEATHERALL DJ. (1985). A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature* **317:** 542-544.

REEDERS ST, BREUNING MH, RYYNANEN MA, WRIGHT AF, DAVIES KE, KING AW, WATSON ML, and WEATHERALL DJ. (1987). A study of genetic linkage heterogeneity in adult polycystic kidney disease. *Hum. Genet.* **76:** 348-351.

REEDERS ST, KEITH T, GREEN P, GERMINO GG, BARTON NJ, LEHMANN OJ, BROWN VA, PHIPPS P, MORGAN J, BEAR JC, and PARFREY P. (1988). Regional localisation of the autosomal dominant polycystic kidney disease locus. *Genomics* **3:** 150-155.

RIGBY PWJ, DIECKMANN M, RHODES C, and BERG P. (1977). Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* **113:** 237-251.

RILEY JH, OGILVIE D, and ANAND R. (1992). Construction, characterisation and screening of YAC libraries. In Anand R. (ed): *Techniques for the Analysis of Complex Genomes.* Academic Press. p59-79.

ROCCO MV, NEILSON EG, HOYER JR, and ZIYADEH FN. (1992). Attenuated expression of epithelial cell adhesion molecules in murine polycystic kidney disease. *Am. J. Physiol.* **262 (4 Pt 2):** F679-F686.

ROMEO G, DEVOTO M, COSTA G, RONCUZZI L, CATIZONE L, ZUCCHELLI P, GERMINO GG, KEITH T, WEATHERALL DJ and REEDERS ST. (1988). A second genetic locus for autosomal dominant polycystic kidney disease. *Lancet* **ii:** 8-11.

ROSENTHAL A, MacKINNON RN, and JONES DSC. (1991). PCR walking from microdissection clone M54 identifies three exons from the human gene for the neural cell adhesion molecule L1 (CAM-L1). *Nucleic Acids Res.* **19:** 5395-5401.

ROSS MT, NIZETIC D, NGUYEN C, KNIGHTS C, VATCHEVA R, BURDEN N, DOUGLAS C, ZEHETNER G, WARD DC, BALDINI A, and LEHRACH H. (1992). Selection of a human chromosome 21 enriched YAC sub-library using a chromosome-specific composite probe. *Nature Genetics* **1:** 284-290.

SARIS JJ, BREUNING MH, DAUWERSE HG, SNIJDEWINT FGM, TOP B, FODDE R, and VAN OMMEN GJB. (1990). Rapid detection of polymorphism near gene for adult polycystic kidney disease. *Lancet* **i:** 1102-1103.

SEDMAN A, BELL P, MANCO-JOHNSON M, SCHRIER R, WARADY BA, HEARD EO, BUTLER-SIMON N, and GABOW P. (1987). Autosomal dominant polycystic kidney disease in childhood: a longitudinal study. *Kidney Int.* **31:** 1000-1005.

SHERRINGTON R, MELMER G, DIXON M, CURTIS D, MANKOO B, KALSI G, and GURLING H. (1991). Linkage disequilibrium between two highly polymorphic microsatellites. *Am. J. Hum. Genet.* **49:** 966-971.

SNELL RG, LAZAROU LP, YOUNGMAN S, QUARRELL OWJ, WASMUTH JJ, SHAW DJ, and HARPER PS. (1989). Linkage disequilibrium in Huntington's disease: an improved localisation for the gene. *J. Med. Genet.* **26:** 673-675.

SOMLO S, GERMINO GG, WIRTH B, WEINSTAT-SASLOW D, BARTON N, GILLESPIE GAJ, FRISCHAUF A-M, and REEDERS ST. (1992a). The molecular genetics of autosomal dominant polycystic kidney disease of the PKD1 type. In Breuning MH, Devoto M, and Romeo G (eds): *Polycystic kidney disease. Contrib. Nephrol.* Basel, Karger **97:** 101-109.

SOMLO S, WIRTH B, GERMINO GG, WEINSTAT-SASLOW D, GILLESPIE GAJ, HIMMELBAUER H, STEEVENS L, COUKE P, WILLEMS P, BACHNER L, COTO E, LOPEZ-LARREA C, PERAL B, SAN MILLAN JL, SARIS JJ, BREUNING MH, FRISCHAUF A-M, and REEDERS ST. (1992b). Fine genetic localisation of the gene for autosomal dominant polycystic kidney disease (PKD1) with respect to physically mapped markers. *Genomics* **13:** 152-158.

TADA S, YAMAGISHI J, KOBAYASHI H, HATA Y, and KOBARI T. (1983). The incidence of simple renal cyst by computed tomography. *Clin. Radiol.* **34:** 437-439.

TAKAHASHI H, UEYAMA Y, HIBINO T, KUWAHARA Y, SUZUKI S, HIOKI K, and TAMAOKI N. (1986). A new mouse model of genetically transmitted polycystic kidney disease. *J. Urol.* **135:** 1280-1283.

THE HUNTINGTON'S DISEASE COLLABORATIVE RESEARCH GROUP. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's Disease chromosomes. *Cell* **72:** 971-983.

THEILMANN J, KANANI S, SHIANG R, ROBBINS C, QUARRELL O, HUGGINS M, HEDRICK A, WEBER B, COLLINS C, WASMUTH JJ, BUETOW KH, MURRAY JC, and HAYDEN MR. (1989). Non-random association between alleles detected at D4S95 and D4S98 and the Huntington's Disease gene. *J. Med. Genet.* **26:** 676-681.

THOMPSON EA, DEEB S, WALKER D, and MOTULSKY AG. (1988). The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. *Am. J. Hum. Genet.* **42:** 113-124.

TORRES VE, DONOVAN KA, SCICLI G, HOLLEY KE, THIBODEAU SN, CARRETERO OA, INAGAMI T, McATEER JA, and JOHNSON CM. (1992). Synthesis of renin by tubulocystic epithelium in autosomal dominant polycystic kidney disease. *Kidney Int.* **42:** 364-373.

TRUDEL M, D'AGATI V, and COSTANTINI F. (1991). c-*myc* as an inducer of polycystic kidney disease in transgenic mice. *Kidney Int.* **39:** 665-671.

TURCO A, PEISSEL B, GAMMARO L, MASCHIO G, and PIGNATTI PF. (1991). Linkage analysis for the diagnosis of autosomal dominant polycystic kidney disease, and for the determination of genetic heterogeneity in Italian families. *Clin. Genet.* **40:** 287-297.

WATSON ML, MACNICOL AM, ALLAN PL, and WRIGHT AF. (1992). Effects of angiotensin converting enzyme inhibition in adult polycystic kidney disease. *Kidney Int.* **41:** 206-210.

WATSON ML, MACNICOL AM, and WRIGHT AF. (1990). Adult polycystic kidney disease: many advances in diagnosis, assessment and counselling. *Br. Med. J.* **300:** 62-63.

WATSON ML, WRIGHT AF, MACNICOL AM, ALLAN PL, CLAYTON JF, DEMPSTER M JEREMIAH SJ, CORNEY G and HOPKINSON DA. (1987). Studies of genetic linkage between adult polycystic kidney disease and three markers on chromosome 16. *J. Med. Genet.* **24:** 457-461.

WEISSENBACH J, GYAPAY G, DIB C, VIGNAL A, MORISSETTE J, MILLASSEAU P, VAYSSEIX G, and LATHROP M. (1992). A second-generation linkage map of the human genome. *Nature* **359:** 794-801.

WERDER AA, AMOS MA, NIELSEN AH, and WOLFE GH. (1984). Comparative effects of germfree and ambient environments on the development of cystic kidney disease in CFWwd mice. *J. Lab. Clin. Med.* **103:** 399-407.

WIEBERS DO, and TORRES VE. (1992). Screening for unruptured intracranial aneurysms in autosomal dominant polycystic kidney disease. [Editorial] *N. Engl. J. Med.* **327:** 953-955.

WILKIE AOM, HIGGS DR, RACK KA, BUCKLE VJ, SPURR NK, FISCHEL-GHODSIAN N, CECCHERINI I, BROWN WRA, and HARRIS PC. (1991). Stable length polymorphism of up to 260 kb at the tip of the short arm of human chromosome 16. *Cell* **64:** 595-606.

WILSON PD, HRENIUK D, and GABOW PA. (1992). Abnormal extracellular matrix and excessive growth of human adult polycystic kidney disease epithelia. *J. Cell. Physiol.* **150:** 360-369.

WILSON PD, SCHRIER RW, BRECKON RD, and GABOW PA. (1986). A new method for studying human polycystic kidney disease epithelia in culture. *Kidney Int.* **30:** 371-378.

WILSON PD, and SHERWOOD AC. (1991). Tubulocystic epithelium. *Kidney Int.* **39:** 450-463.

WILSON PD, SHERWOOD AC, PALLA K, DU J, WATSON R, and NORMAN JT. (1991). Reversed polarity of Na^+-K^+-ATPase: mislocation to apical plasma membranes in polycystic kidney disease epithelia. *Am. J. Physiol.* **260:** F420-F430.

WRIGHT AF, TEAGUE PW, POUND SE, PIGNATELLI PM, MACNICOL AM, CAROTHERS AD, De MEY RJ, ALLAN PL, and WATSON ML. (1993). A study of genetic linkage heterogeneity in 35 adult-onset polycystic kidney disease families. *Hum. Genet.* **90:** 569-571.

ZERRES K, VÖLPEL MC, and WEIß. (1984). Cystic kidneys. Genetics, pathologic anatomy, clinical picture, and prenatal diagnosis. *Hum. Genet.* **68:** 104-135.

APPENDIX ONE

Forward sequence of VK5/*Msp*I subclone

5' **TGG TAC CGG GCC CCC CCT CGA GGT CGA CGG TAT CG:**G CGT CAA CAC GGG ATA ATA CCG CGC CAC ATA GCA GAA CTT TAA AAG TGC TCA TCA TTG GAA AAC GTT CTT CGG GGC GAA AAC TCT CAA GGA TCT TAC CGC TGT TGA GAT CCA GTT CGA TGT AAC CCA CTC GTG CAC CCA ACT GAT CTT CAG CAT CTT TTA CTT TCA CCA GCG TTT CTG GGT GAG CAA AAA CAG GAA GGC AAA ATG CCG CAA AAA AGG GAA TAA GGG CGA CAC GGA AAT 3'

Sequence obtained using the M13 universal primer with the VK5/*Msp*I subclone. The pBluescript SK vector sequence is shown in bold type. The insert sequence shows 100% homology with pBR322 sequence, bases 3905-4142. The GC content of the sequence is 48%.

Reverse sequence of VK5/*Msp*I subclone

5' **GGC AAC AAA AGC TGG TAC CGG GCC CCC CCT CGA GGT CGA CGG TAT CG:**G TCA GAG GCC ATG CTC ACG TGT CCT GAG GTC CCC GAT GAC AAG CCA CAA TGC CAT GGC ACA GTG GCC CAT GCA GCA CTC AGG TGG CTT CCG TGG AAG CTG CTG AAG AAG GCA TGC CCT GCA CAG ACC CAT CAG ACC CCA CCC GCC GCA GCT TGC CAA AAC CAG AGA GAA GGC AAC GAG GGC TGG CAG CAC CTT CCA GCA TGG GCG GGA AGG GGA GGC 3'

Sequence obtained using the M13 reverse primer with the VK5/*Msp*I subclone. The pBluescript SK vector sequence is shown in bold type. The VK5 sequence has a GC content of 63%. The VR1 primer sequence is underlined.

VK5 sequence from the pBR322 *Eco*RI site

5' GGA CGC TCT CGT GTC GTC ATG TCG AGC CAC AGG TGC TCT TGG GAT TTC AGG AAC ATG ACT CAT TTT ATT TTG GAA CTT AGA AAT AGA ATA TTT GGA GGC GAA CGA GGA AGC AGA AAA CAG CAG CTG TTC GGG CCG CGT GTT TTG CTT TTG TGG ATG TGC TCA GGA TTG CTT TGG GCC TCG TGC 3'

VK5 sequence obtained using a 15 mer pBR322 *Eco*RI site clockwise primer (5' GTA TCA CGA GGC CCT 3'). The primer was very close to the cloning site and no vector sequence was obtained. The GC content of the sequence is 49%. The VF1 primer sequence is underlined.

VK5 sequence from the pBR322 *Bam*HI site

5' GCC CCT CGG CCC CCT GCC CTC CCC ACC CTC CAA TCT CCT CAA AAC TCT TCC TCA AGG TAC CCT CCC CAC TTC CCT CAC CAG TGT CCA GGA ACC CCC TGA GGC CAC CCC CAA GTC GGC CTG GGC TCC CCT AGG ACA GGC TCA GGC GCT GGC AAT GCT GCT GAG GTG AAC TCC AAA ATG AGC ATG GCA GCT CAG AGA CCA GG 3'

VK5 sequence obtained using a 16 mer pBR322 *Bam*HI site counter-clockwise primer (5' ATG CGT CCG GCG TAG A 3'). The primer was very close to the cloning site, and no vector sequence was obtained. The GC content of the sequence is 62%.

Forward sequence of VK5/*Kpn*I subclone

5' **CAC CGC GGT GGC GGC CGC TCT AGA ACT AGT GGA TCC CCC GGG CTG CAG GAA TTC GAT ATC AAG CTT ATC GAT ACC GTC GAC CTC GAG GGG GGG CCC GGT AC:**C CAC CCC TGC ACT CAG GGA TGG CCC AGA CCC CCA CCA CGA CAA CTG GAC CCC CAT GTC CCT GCA CTC AGG GAT GGC CCA GAC CCC TAT GAC AGC CAG GCC TCA GCA TCC CAA GTC CTC ACT CAC ACA GCA GAG GCT GGG CCC GGG GAA CAG AGG ACC CTG GGT CCC TGA GAC ACA GCC CTC AGT GGC ATT CCT AAG AGC ATG GCC CT 3'

Sequence obtained using the M13 universal primer with the VK5/*Kpn*I subclone. The pBluescript KS vector sequence is shown in bold type. A 28 bp repeat sequence is underlined. The GC content of the sequence is 64%.

Reverse sequence of VK5/*Kpn*I subclone

5' **AAA GGG AAC AAA AGC TGG GTA C:**CC TTA GCT GCT TGG GCC CAT GCC GGC ACC CCT GTG GCC GCC TCA CTG GGC TGG GTC CAC TGC TGC CCA CTG GGT CAG CCG CCA GCT ATA CAG GAG CAG CAG GTA GCT GGG GCT GCT CTC CCG GCC GTT GGT CCA ACT GCC CAG GTC TTA GAG AGG CCA AAG TGT GAG CCT CGT CCA GCT GCT TTT TGG CTT GGT CAT CTG CTG CTG TCA TGC TTA TTG TGA TTT GCG 3'

Sequence obtained using the M13 reverse primer with the VK5/*Kpn*I subclone. The pBluescript KS vector sequence is shown in bold type. The GC content of the sequence is 62%.

Sequence of AJ1 cDNA from which the ATPL primers were designed

CACAGAATAT TATGTAAAGA CCACCCCTCC TCATTCCAGA ACGAACAGCC

TGACACATCT GCAGCGCCGC CCGCCCCCAG TAGTTGGTCT TGTACATGCG
B928 C18

CAGTATCCTA GTGCCCATCG TCTGTTTCCC CGCCTTGCCC CCGCCCGCCC

CGTGCCGTGG ACATCTGGGC CCACTCATCG CCCCTCCAGG CCCCCGGCGC

CCCACCCCCT AAAGTGCTCT AGTATGCGGA TGATTTAGAA TTGTCATTTC

↓ *Bgl*II
TCTTTACTGG ATGTTTATTA TTAAAGATCT CGCCTGTTCC TGCGTCTGCG

GAGCCGCCCT TGTCTCCCAG CTATCTATAA CCTTAGCTTG TGTGTCGCCT
ACAGCGGA

TGTGGGTTCC TGTTGCTGAG ACTTTTCCTG GATGGAGCCG CCCTCACCGC
ACACCCAAGG ACAA
C19

GCCCGTGGCC CTGCGCGGAG CTGTGTCCAA TAAAGTTCTT GCT
CGCGCCTC GACACAGGTT ATT
B927

Sequence of the 3' untranslated region of AJ1 (ATPL) cDNA, with the nucleotides shown 5' to 3'. The four primer sequences tested are underlined and the position of the unique *Bgl*II site marked.

APPENDIX TWO

PUBLISHED MATERIAL

Hum Genet (1993) 90:569–571

human genetics

A study of genetic linkage heterogeneity in 35 adult-onset polycystic kidney disease families

Alan F. Wright[1], Peter W. Teague[1], Susan E. Pound[1,2], Patricia M. Pignatelli[1], Anne M. Macnicol[2], Andrew D. Carothers[1], Rhona J. De Mey[1], Paul L. Allan[2], Michael L. Watson[2]

[1] MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK
[2] Department of Medicine, University of Edinburgh and Royal Infirmary of Edinburgh, Lauriston Place, Edinburgh, EH3 9YW, UK

Received: 15 September 1991 / Revised: 1 July 1992

Abstract. A genetic heterogeneity analysis of 35 kindreds with adult-onset polycystic kidney disease (ADPKD) was carried out using the D16S85, D16S84, D16S125 and D16S94 loci that are closely linked to the PKD1 locus on chromosome 16. The results show that the likelihood of two ADPKD loci is 2,514.9 times greater than for a single locus ($P < 0.0001$). The maximum likelihood lod score is 27.38 under heterogeneity with PKD1 lying 4.9 c*M* proximal to D16S85 (in males). At least 3% of kindreds are unlinked to PKD1, since the 95% confidence limits of alpha, the proportion of families linked to PKD1, are 0.54–0.97. Only 2 out of 35 kindreds (5.7%) show statistically significant evidence of non-linkage to PKD1, with conditional probabilities of 0.987 and 0.993 that the disease locus is unlinked. This confirms the existence of a small subgroup of ADPKD kindreds that are unlinked to PKD1 and provides a firm basis for genetic counselling of this population on the basis of DNA probes.

Introduction

The first study of genetic heterogeneity in adult-onset polycystic kidney disease (ADPKD) was reported by Reeders et al. (1987) who found no evidence to support heterogeneity in a sample of 27 northern European families from four countries. Subsequently, two reports have been published describing families that appear unlinked to this region of chromosome 16 (Romeo et al. 1988; Kimberling et al. 1988). We have extended our initial heterogeneity study of 10 ADPKD families, included in the report of Reeders et al. (1987) to a total of 35 families from central Scotland to clarify the extent of genetic heterogeneity in this population.

Materials and methods

Ascertainment and diagnosis of family members

Families were ascertained through the Medical Renal Unit, Royal Infirmary of Edinburgh and the Renal Unit, Western Infirmary, Glasgow. In all, 343 members of 35 ADPKD families were admitted to the study. Diagnostic criteria included the presence of two or more cysts greater than 0.5 cm in diameter in one kidney and at least one such cyst in the other kidney, together with a family history of ADPKD. Ultrasound examinations were carried out by experienced sonologists, one in each centre, as described previously (Watson et al. 1987; Pignatelli et al. 1992). Pedigrees were drawn up on the basis of interviews with patients and relatives and checked by consultation with the centralised Register for Births, Marriages and Deaths for Scotland. After obtaining informed consent, as approved by the local ethical committee, 30–50 ml EDTA anticoagulated blood was removed from family members and frozen at –70°C prior to extraction of DNA.

DNA extraction and analysis

DNA was extracted and analysed for markers as described by Pignatelli et al. (1992) except that in addition to D16S85, D16S84 and D16S94, the D16S125 (26-6) locus (Breuning et al. 1990) was analysed.

Linkage and heterogeneity analysis

Five-point linkage analyses were carried out using the LINKMAP subroutine of the LINKAGE program package version 5.03 (Lathrop et al. 1984). Recombination fractions were estimated at intervals of 0.01 between 0 and 0.4. Mutation rates were set at 2×10^{-6} per ADPKD locus per gamete for both sexes and the ADPKD gene frequency at 0.0005. The program was run using different fixed ratios of recombination in males to females, which identified a broad likelihood peak ratio in the region of 5 ($\theta m : \theta f$). This value was therefore used for the final analysis. The resulting family lod scores were then analysed using the HOMOG program package (Ott 1985) for linkage heterogeneity. The probability of detecting a PKD1 gene carrier by ultrasound analysis was taken to be 0.64, 0.92 and 1.00 for ages 0–10 years, 10–30 years and over 30 years respectively (Bear et al. 1991).

Results

Multipoint heterogeneity analysis

Multipoint linkage analysis was carried out using probe data from the D16S85, D16S84, D16S125 and D16S94 loci. The

Correspondence to: A.Wright

D16S84 locus is more closely linked to PKD1 than is D16S85 and both are thought to be distal to PKD1, whereas D16S125 and D16S94 are probably proximal (Gillespie et al. 1991). The five-point lod scores were analysed for linkage heterogeneity using the HOMOG program package. The results are shown in Table 1.

The log (ln) likelihood on the assumption of a single ADPKD (linkage homogeneity) was 55.22 compared with 63.05 (lod score 27.38) on the assumption of one linked and one unlinked ADPKD locus (linkage heterogeneity). The likelihood ratio is 2,514.9, which is statistically significant (chi-square 15.66; $P < 0.0001$). The maximum likelihood value of α, the proportion of families linked, was found to be 0.81 with approximate 95% confidence limits of 0.54–0.97, showing that at least 3% of families are unlinked. The recombination fractions separating ADPKD from the anchor locus D16S85 were 0.049 (males) and 0.010 (females) under linkage heterogeneity. Linkage heterogeneity is therefore present in this sample of families.

Table 1. Results of heterogeneity analyses in 35 adult-onset polycystic kidney disease (ADPKD) kindreds. Multipoint lod scores for ADPKD, D16S85, D16S84, D16S125 and D16S94 were analysed with the HOMOG program. The maximum likelihood values of the proportion of families linked, α, and of the recombination fractions, measured proximally from D16S85, are shown for males (θ_m) and females (θ_f) separately with approx. 95% confidence limits (α, θ) in parentheses

	Linkage heterogeneity	Linkage homogeneity
α (proportion of families linked)	0.81 (0.54 –0.97)	1.00
θ_m	0.049 (0.025–0.059)	0.047
θ_f	0.010 (0.005–0.012)	0.009
Log (ln) likelihood	63.05	55.22
P value	<0.0001 (chi-square = 15.66)	
Likelihood ratio	2,514.9	

Table 2. Conditional probabilities of linkage to PKD1 are shown for each of 35 ADPKD kindreds studied. The results were obtained using the HOMOG program

Family number	Conditional probability of linkage to PKD1	Family number	Conditional probability of linkage to PKD1
PK1	0.784	PK30	0.997
PK2	1.000	PK31	0.795
PK3	0.995	PK32	0.967
PK4	1.000	PK35	0.895
PK5	0.932	PK36	0.930
PK6	0.999	PK37	0.751
PK7	1.000	PK38	0.886
PK8	0.747	PK39	0.780
PK11	0.999	PK41	0.860
PK17	0.999	PK42	0.998
PK18	0.999	PK45	0.961
PK19	0.871	PK49	0.958
PK24	0.881	PK50	0.608
PK25	0.460	PK52	0.013
PK26	0.895	PK53	0.007
PK27	0.574	PK55	0.602
PK28	0.982	PK56	0.886
PK29	0.289		

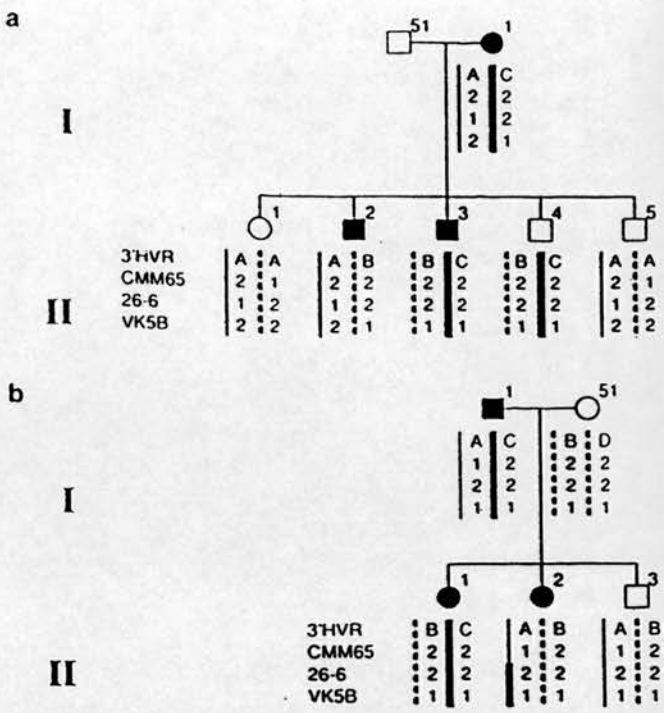


Fig. 1. a Haplotype analysis of family PK53, in which there are either two or three double recombinants with ADPKD, depending on the linkage phase, on the assumption of a PKD1 locus. There is no evidence of recombination between markers. Affected individuals are shown as *shaded boxes* (male) or *circles* (female). The unaffected individuals II-1, II-4 and II-5 are aged 39, 32 and 31 years respectively. Possible haplotypes are shown, with results using the following probes: D16S85 (3'HVR), D16S84 (CMM65), D16S125 (26-6) and D16S94 (VK5B). **b** Haplotype analysis of family PK52 in which a single recombination (in II-2) would be consistent with a PKD1 locus although no exchange of flanking markers has been demonstrated

Families that appear unlinked to PKD1

Only 2 small families out of 35 analysed show statistically significant evidence of non-linkage to PKD1 (Table 2), with probabilities of 0.013 (PK52) and 0.007 (PK53) that they are linked to PKD1, conditional on the maximum likelihood values of α (0.81) and θ_m (0.049). One family (Fig. 1a) (PK53) is informative with flanking markers D16S85 and D16S84 distally and D16S125 proximally and would require either two or three double recombinants out of five meioses to be consistent with a PKD1 locus. No exchange of flanking markers is seen. This family therefore does appear unlinked from haplotype analysis. The other family (PK52) (Fig. 1b) is informative for D16S85 and D16S84 distally but with neither proximal marker, so that only a single recombinant (with D16S85 and D16S84) out of three meioses needs to be invoked to be consistent with a PKD1 locus. Further loci are currently being screened. The evidence for non-linkage in this family is therefore ambiguous.

Discussion

Two families have been described in the literature that provide clear evidence of non-linkage to the alpha globin (HBA) region on chromosome 16 (Romeo et al. 1988; Kimberling et al. 1988) and several others have been reported at meetings. By all accounts, the frequency of unlinked families is low in the experience of most laboratories, although only two formal heterogeneity analyses have been published in full (Reeders et al. 1987, Mandich et al. 1990).

Reeders et al. (1987) found no evidence of linkage heterogeneity in 27 ADPKD families from the United Kingdom, the Netherlands and Finland using D16S85 data alone. Mandich et al. (1990) studied 29 Italian families with ADPKD and also found no significant evidence of heterogeneity using 2-point analysis with D16S85 or D16S80 although the peak value of α, the proportion of families linked, was 0.92 (lod score 12.97). In the present study, the likelihood of two loci is 2,514.9 times greater than for a single locus although only one small kindred (Fig. 1a) provides convincing evidence of non-linkage to PKD1.

Phenotypic differences between PKD1 and ADPKD families unlinked to PKD1 (PKD2) have been described. Bear et al. (1989) reported the presence of a milder phenotype in PKD2 families with delayed appearance of cysts and hypertension and a later age at which dialysis becomes necessary. The phenotype in family PK53 is consistent with this, although it does not appear uniformly mild within the family. Family PK52 shows the typical PKD1 phenotype. Both families are of Scottish ancestry.

What are the implications of these results for presymptomatic diagnosis of ADPKD? Although the available evidence suggests that there has been little demand for this in families (Watson et al. 1990), the low level of heterogeneity observed is likely to have a relatively small effect on the accuracy of genetic diagnosis. Caution should be exercised however if recombination is observed in a family in the absence of flanking marker exchange.

Acknowledgements. The study was generously supported by the National Kidney Research Fund and The Wellcome Trust whose support is gratefully acknowledged. We would also like to thank Drs. B. J. R. Junor, D. Briggs and S. Roger of the Renal Unit, Western Infirmary Glasgow for kindly providing access to their patients, and Mrs. E. Kay for typing the manuscript.

References

Bear JC, Parfray PS, Morgan JM, Cramer BC, McManamon PJ, et al (1989) Autosomal dominant polycystic kidney disease: ultrasonographic detection and prognosis of PKD1 and PKD2 forms. Am J Hum Genet 45 : [Suppl] A39

Bear JC, Parfray PS, Morgan JM, Martin CJ, Cramer BC (1991) Autosomal dominant polycystic kidney disease: new information for genetic counselling. Am J Hum Genet 49 : [Suppl] A226

Breuning MH, Snijdewint FG, Smits JR, Dauwerse JG, Saris JJ, Ommen GJ van (1990) A *Taq*I polymorphism identified by 26-6 (D16S125) proximal to the locus affecting adult polycystic kidney disease (PKD1) on chromosome 16. Nucleic Acids Res 18 : 3106

Gillespie GAJ, Somlo S, Germino GG, Weinstat-Saslow D, Reeders ST (1991) CpG island in the region of an autosomal dominant polycystic kidney disease locus defines the 5′ end of a gene encoding a putative proton channel. Proc Natl Acad Sci USA 88 : 4289–4293

Kimberling WJ, Fain PR, Kenyon JB, Goldgar D, Sujansky E, Gabow PA (1988) Linkage heterogeneity of autosomal dominant polycystic kidney disease. New Engl J Med 319 : 913–918

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81 : 3443–3446

Mandich P, Restagno G, Novelli G, Bellone E, Potenza L (1990) Autosomal dominant polycystic kidney disease: a linkage evaluation of heterogeneity in Italy. Am J Med Genet 35 : 579–581

Ott J (1985) Analysis of human linkage. Johns Hopkins, Baltimore

Pignatelli PM, Pound SE, Carothers AD, Macnicol AM, Allan PL, Watson ML, Wright AF (1992) Multipoint mapping of adult-onset polycystic kidney disease (PKD1) on chromosome 16. J Med Genet 29 : 638–641

Reeders ST, Breuning MH, Ryynanen MA, Wright AF, Davies KE, King AW, Watson ML, Weatherall DJ (1987) A study of genetic linkage heterogeneity in adult polycystic kidney disease. Hum Genet 76 : 348–351

Romeo G, Costa G, Catizone L, Germino GG, Weatherall DJ, Devoto M, Roncuzzi L, Zucchelli P, Keith T, Reeders ST (1988) A second genetic locus for autosomal dominant polycystic kidney disease. Lancet ii : 7–10

Watson ML, Wright AF, Macnicol AM, Allan PL, Clayton JF, Dempster M, Jeremiah SJ, Corney G, Hopkinson DA (1987) Studies of genetic linkage between adult polycystic kidney disease and three markers on chromosome 16. J Med Genet 24 : 457–461

Watson ML, Macnicol AM, Wright AF (1990) Adult polycystic kidney disease. Br Med J 300 : 62

J Med Genet 1992; 29: 638–6

Multipoint mapping of adult onset polycystic kidney disease (PKD1) on chromosome 16

P M Pignatelli, S E Pound, A D Carothers, A M Macnicol, P L Allan, M L Watson, A F Wright

Abstract

Analysis of genetic linkage data in 33 adult onset polycystic kidney (ADPKD) families was carried out using probes for the *D16S85, D16S84,* and *D16S94* loci. The data set of 33 families shows no evidence of genetic heterogeneity since one unlinked family was previously excluded. Two point linkage analysis showed maximum likelihood values of the recombination fraction of 0·07 for *ADPKD* and *D16S85* (lod score 18·78), 0·02 for *ADPKD* and *D16S84* (lod score 7·55), and 0·00 for *ADPKD* and *D16S94* (lod score 6·73). Multipoint analysis showed a maximum likelihood order of tel–*D16S85*–0·06–*D16S84*–0·02–(*PKD1, D16S94*)–cen with a multipoint lod score of 32·16. Analysis of rare recombinants lying close to *PKD1* gave results consistent with this order.

(*J Med Genet* 1992;29:638–41)

Adult onset polycystic kidney disease (ADPKD) is an autosomal dominant condition characterised by progressive formation and enlargement of multiple cysts in the kidney and other organs leading to deterioration in renal function and the development of hypertension in middle life. End stage renal disease is a common outcome although renal replacement programmes have improved the prognosis considerably.[1] Premature deaths from the complications of renal failure and rupture of intracranial aneurysms are not uncommon.[2] ADPKD is one of the commonest genetic diseases in man affecting 1 per 1000 of the general population.[3] The pathophysiological basis is not well understood although the earliest changes appear to include hyperplasia of tubular epithelial cells and focal microcyst formation arising in any portion of the nephron from glomerulus to collecting tubules.[4] Abnormalities in tubular basement membrane, sodium pump orientation, and sensitivity to growth factors have recently been identified in ADPKD kidneys and cultured tubular epithelial cells.[5]

Linkage was first shown between *ADPKD* and *D16S85* close to α globin in chromosomal region 16p13 by Reeders *et al*.[6] This localisation has since been amply confirmed by several other groups.[7–9] The orientation of the *ADPKD* locus relative to *D16S85*, afterwards localised to band 16p13.3,[10,11] was shown in a multipoint linkage analysis[11] which indicated that the odds in favour of *PKD1* lying proximal to *D16S85* were >10 000:1 compared with a distal location. This group also showed that the *D16S63* and *D16S45* loci were locate proximal to both *PKD1* and *D16S85* by combination of linkage and somatic cell hybri analyses.[11] Breuning *et al*[9] also showed that th disease locus is flanked by *D16S80* proximal and *D16S85* distally.

A previous analysis of 27 ADPKD famili from four countries failed to show evidence genetic heterogeneity.[12] However, two repor subsequently appeared describing families th failed to show linkage to the α globin region chromosome 16.[13,14] Since accurate localisatio of the gene first requires a genetically hom geneous sample, we extended our initial stud of 10 ADPKD families, included in the stud of Reeders *et al*,[12] to a total of 34 families. Th results of this study are described elsewhere and show that 81 to 83% of these families sho linkage to this region of chromosome 16. How ever, only one family (PK53) showed cle evidence of non-linkage to *PKD1*[15] and w therefore removed from the data set for th purpose of this multipoint analysis. We no report the results of the analysis of the remai ing 33 families using *D16S85, D16S84,* a *D16S94*.

Materials and methods

ASCERTAINMENT AND DIAGNOSIS OF FAMILY MEMBERS

Families were ascertained through the Medic Renal Unit, Royal Infirmary of Edinburgh a the Renal Unit, Western Infirmary, Glasgo Diagnostic criteria included a family histo of ADPKD (more than one affected membe and ultrasound findings of two or more cy greater than 0·5 cm in diameter in one kidn and at least one such cyst in the contralate kidney.[7] Pedigrees were drawn up after int views with patients and were checked agai the centralised Register for Births, Marriag and Deaths for Scotland. Ethical approval w obtained and, after obtaining informed co sent, 30 to 50 ml EDTA anticoagulated blo were drawn and frozen at −70°C before DN extraction. Thirty-four probands and 3 family members were initially ascertained a their relatives sampled for biochemical a DNA analysis and examined clinically as d scribed previously,[7] after which ultrasou examinations were arranged and carried out experienced sonologists, one in each centre

A previous study of linkage heterogenei identified one family (PK53) contribut disproportionately to the observed hete geneity and so was excluded from this stu This family showed a conditional probabi of 0·003 of being linked to the *D16S*

MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU.
P M Pignatelli
S E Pound
A D Carothers
A F Wright

Department of Medicine, Royal Infirmary of Edinburgh, Lauriston Place, Edinburgh EH3 9YW.
S E Pound
A M Macnicol
P L Allan
M L Watson

Correspondence to Dr Wright.

Received 16 September 1991. Revised version accepted 2 March 1992.

D16S84, and *D16S94* loci. None of the other families showed good evidence of non-linkage to this region, so this set of 33 families was used in the study.

DNA EXTRACTION AND ANALYSIS

DNA was extracted by the method of Kunkel *et al.*[16] The following loci were used to type all family members.

(1) *D16S85* (3'HVR) is a hypervariable locus described by Jarman *et al*[17] and located in 16p13.3, which detects multiple alleles with several restriction enzymes. DNA samples were digested with the restriction enzyme *Pvu*II and allele frequencies were each set at 0·1.

(2) *D16S84* (pCMM65) was isolated by Nakamura *et al*[18] and detects two alleles of size 3·5 kb (A1) and 2·1 kb (A2) with frequencies of 0·40 and 0·60 with the enzyme *Pvu*II. It is localised to chromosomal region 16p.

(3) *D16S94* (pVK5B) was isolated by Hyland *et al*[19] and detects alleles of 1·6 kb (A1) and 1·3 kb (A2) with the enzyme *MspI* with frequencies of 0·55 and 0·45. It is localised to chromosomal region 16p13.3.

(4) *D16S45* (CRI-090) was isolated by Donis-Keller *et al*[20] and detects two alleles (20 kb and 13 kb) with frequencies of 0·47 and 0·53 with *Eco*RI. This probe was only used to probe families containing key recombinants. It is localised to chromosomal region 16pter-p13.

Probes were used without purification of inserts and labelled with ^{32}P-dCTP by random priming.[21] Unincorporated counts were separated by gel filtration. DNA from the patients was digested to completion with the above enzymes and separated by electrophoresis on 0·8% agarose gels in 1 × TBE buffer. Transfers were carried out by the method of Southern[22] using nylon (Nytran) filters. Filters were prehybridised for at least four hours in 5 × Denhardt's solution/4 × SSC/10% dextran sulphate/0·1% sodium pyrophosphate/1% sodium dodecylsulphate (SDS)/0·1 mg ml^{-1} denatured salmon sperm DNA. Filters were then hybridised in the same mixture containing 1 to 2 × 10^6 cpm ml^{-1} of labelled probe (2·5 ng ml^{-1}). After overnight hybridisation at 68°C, filters were washed in 2 × SSC/1% SDS down to a final stringency of 0·1 × SSC/1% SDS at 68°C. Filters were used to expose Kodak XAR-5 films in cassettes containing double intensifier screens for one to 14 days at −70°C.

LINKAGE ANALYSIS

Linkage was analysed using the LINKAGE program package version 5·03.[23] The ADPKD gene frequency was set at 0·0005 and male and female mutation rates assumed to be equal at 5 × 10^{-5} per locus per gamete. The female:male recombination ratio was assumed to be constant and was found to maximise the likelihood at a ratio of 0·2. This value was used for LINKMAP runs. MLINK was unable to support a fixed ratio of recombination in the two sexes, so sex averaged values were obtained. Two point linkage analyses were obtained using the MLINK subroutine. Multipoint analyses were run using LINKMAP with a fixed order of marker loci as follows: *D16S85*–0·06–*D16S84*–0·02–*D16S94*. The probabilities of ultrasonographic detection in gene carriers were assumed to be 0·22, 0·66, 0·86, and 0·95 during the first four decades of life and were taken to be 1·00 after the age of 40.[24]

Results

TWO POINT LINKAGE ANALYSES

The results of two point linkage analyses are shown in the table. The maximum likelihood value of the recombination fraction (sex averaged) was found to be 0·07 at a lod score of 18·78 for *PKD1-D16S85*. The corresponding values of the recombination fractions were 0.02 at a lod score of 7·55 for *PKD1-D16S84* and 0·00 at a lod score of 6·73 for *PKD1-D16S94*. All three loci show close linkage to *PKD1* and highly significant lod scores. No recombination was found with *D16S94* and only a single definite recombinant with *D16S84*.

MULTIPOINT LINKAGE ANALYSIS

The results of the multipoint analysis is shown in the figure. In a four point analysis with *PKD1, D16S85, D16S84,* and *D16S94*, the maximum likelihood is found to occur at the location *D16S85*–0·06–*D16S84*–0·02–(*PKD1, D16S94*) – cen (order 1) at a peak lod score of 32·16. Since no definite recombination was observed between *PKD1* and *D16S94*, the order of these two loci cannot be determined. This result is consistent with the presence of a single triply informative meiosis in an affected member of family PK52 which is recombinant with both *D16S85* and *D16S84*, although the family has not yet been found to be informative for *D16S94* or other proximal markers. Exchange of flanking markers has therefore yet to be shown. However, in the data set as a whole the likelihood of the order *D16S85*–0·04 – *PKD1* – 0·02 – *D16S84* – 0·02 – *D16S94* – cen (order 2) is only 2·6 times lower than order 1, with *PKD1* proximal to *D16S84*. However, order 1 is at least 2·7 × 10^5 times more likely than with *PKD1* distal to *D16S85*.

A total of 20 recombinants out of 226 informative meioses has been identified in the 33 APKD families using the *D16S85, D16S84, D16S94,* and *D16S45* loci. Only 12 of these are multiply informative and therefore provide information on the localisation of the *PKD1* locus. Five out of eighteen *D16S85* recombinants are also informative for *D16S84*, four of which are non-recombinant with *D16S84*, supporting the proposal that this locus is closer to *PKD1* than *D16S85*.[25] Five out of eighteen *D16S85* recombinants are informative for *D16S94*, each of which is non-recombinant with APKD, consistent with the hypothesis that *PKD1* lies distal or close to *D16S94*.

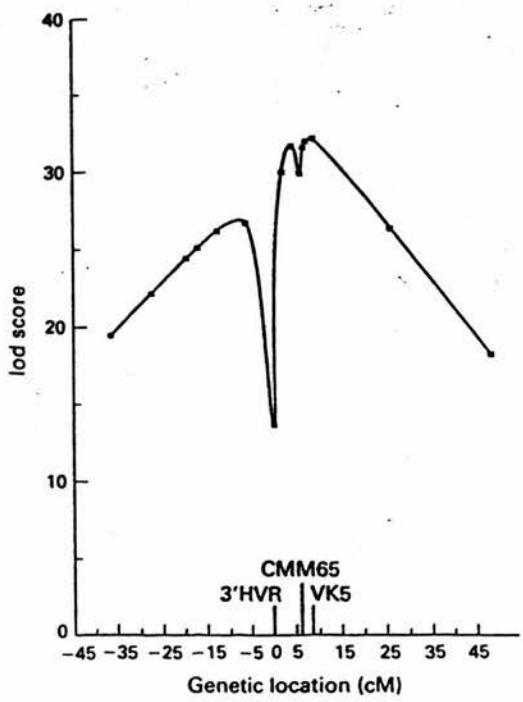
The results of two point linkage analyses of ADPKD and the D16S85, D16S84, and D16S94 loci. Recombination fractions and the corresponding lod scores are shown. The maximum likelihood values of the recombination fractions (θmax) and lod scores (Zmax) are indicated.

Locus	0·00	0·01	0·05	0·10	0·20	0·30	0·40	θmax	Zmax
D16S85 (3'HVR)	2·45	14·82	18·62	18·37	14·53	9·02	3·27	0·07	18·78
D16S84 (CMM65)	6·14	7·42	7·39	6·68	4·75	2·63	0·76	0·02	7·55
D16S94 (VK5B)	6·73	6·56	5·89	5·03	3·31	1·70	0·48	0·00	6·73

Discussion

Previous multipoint analyses of the *PKD1* locus have been described by Reeders *et al*,[11] Germino *et al*,[25] and Breuning *et al*.[9 26] A very substantial body of evidence supports the localisation of *PKD1* proximal to *D16S85*, the most distal marker known on chromosome 16, and distal to the *D16S45, D16S63,* and *D16S80* loci. Further refinement of the map of this region has come from the combined use of linkage and human-rodent hybrid cell lines which have distinguished four subregions by means of *23HA* (GM2324), *N-OH1*, and *CY14* breakpoints.[25 26] These regions separate the most important polymorphic loci into the following four groups, extending distally: (1) *D16S45, D16S63, D16S80* (proximal to *23HA*), (2) *D16S94, D16S125* (proximal to *N-OH1*, distal to *23HA*), (3) *D16S84, D16S259* (proximal to *CY14*, distal to *N-OH1*), (4) *D16S83, D16S21, D16S85* (distal to *CY14*).

The genetic relationship of *PKD1* with loci within groups 2 and 3 is less clear, however. Most of these loci show few if any recombinations with *PKD1*, which on the one hand supports the view that they lie closer to the gene than those in groups 1 and 4, but on the other hand makes ordering difficult. In the absence of ADPKD patients showing chromosomal rearrangements, the genetic definition of closely linked flanking markers is an essential prelude to successfully localising the gene by physical means. Initially, no recombinants were identified with *D16S84*,[25] although at least one has since been reported.[27] Similarly, at least one recombinant has been reported with *D16S125*,[26] which, together with the above, is consistent with a localisation for *PKD1* between *D16S84* and *D16S125*, an area of less than 750 kb as determined by pulsed field gel electrophoresis (PFGE) mapping.[28] However, several potential pitfalls exist which can lead to a misleading genetic assignment, so that it is important to accumulate as much genetic evidence as possible. Firstly, genetic heterogeneity is now known to exist in ADPKD so that genetic mapping should be carried out as far as possible on a genetically homogeneous population. We have attempted to do this by first carrying out a heterogeneity study and then by removal of unlinked families from further analysis.[15] Secondly, false positive[11] and false negative[24] diagnoses can lead to misinterpretation of results. Thirdly, double recombinants, although very rare in such small genetic distances, can occur and may be mimicked by gene conversion events. Fourthly, mistyping can occur, so that it is important that each apparent recombinant is checked, preferably by an independent laboratory. Finally, non-paternity can arise and lead to erroneous interpretation of results, so that this should ideally be checked by genetic fingerprinting.



Results of multipoint analysis of ADPKD and D16S85 (HVR), D16S84 (CMM), and D16S94 (VK5) loci. The multipoint lod scores are plotted against genetic location (based on recombination fractions in males) with reference to the D16S85 (3'HVR) locus at position O. The location of the D16S84 and D16S94 loci are also shown. Distances are given in centiMorgans (cM).

The results presented in this study confirm the tight genetic linkage of *D16S84* in region 3 which shows only a single definite recombinant with *PDK1* (θmax = 0·02, Zmax = 7·55), similarly with *D16S94* in region 2, which shows no definite recombination with *PDK1* (θmax = 0·00, Zmax = 6·73). Multipoint analysis showed the most likely order to be: tel–*D16S85*–*D16S84*–(*PKD1*, *D16S95*)–cen which gave a multipoint lod score of 32·16.

A single recombinant with *D16S84* was also recombinant for *D16S85*, arguing in favour of a location for *PKD1* proximal to both these loci. The recombinant subject is unequivocally affected, so that a false positive diagnosis seems highly unlikely, but the family has so far been uninformative for proximal markers such as *D16S94*. It will therefore be important to show that exchange of flanking markers has occurred in this instance, since the family is not large enough to be unequivocally linked to *PKD1*. Fine mapping these and other rare recombinants in regions 2 and 3, using new polymorphisms identified from these locations will help to confirm the location of *PKD1* and to narrow down the area of search by physical mapping methods.

We would like to acknowledge the generous support of The Wellcome Trust. We would like to thank Drs D R Higgs, Y Nakamura, and

V J Hyland for the use of the *D16S85*, *D16S84*, and *D16S94* probes. We are grateful to Norman Davidson and staff for the artwork, and to Mrs Rhona De Mey for checking pedigrees against the centralised register of births, marriages, and deaths. We would like to thank Drs J D Briggs and B Junor for their cooperation and for access to their patients.

1 Chester AC, Argy WP Jr, Rakowski TA, Schreiner GE. Polycystic kidney disease and chronic hemodialysis. *Clin Nephrol* 1978;10:129–33.
2 Segal AJ, Spataro RF, Barbaric J. Adult polycystic kidney disease: a review of 100 cases. *J Urol* 1977;118:711–13.
3 Dalgaard OZ. Polycystic disease of the kidneys: a follow-up study of 284 patients and their families. *Acta Med Scand (Suppl)* 1957;158:328.
4 Friedman J. Cystic diseases of the kidney. In: Emery AEH, Rimoin DL, eds. *Principles and practice of medical genetics*. Volume I. Edinburgh: Churchill Livingstone, 1983:1002–10.
5 Wilson PD, Sherwood AC, Palla K, Du J, Watson R, Norman JT. Cellular and molecular mechanisms of cyst formation in human autosomal dominant polycystic kidney disease. *Am J Physiol* 1991;260:F420–30.
6 Reeders ST, Breuning MH, Davies KE, *et al*. A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature* 1985;317:542–4.
7 Watson ML, Wright AF, Macnicol AM, *et al*. Studies of genetic linkage between adult polycystic kidney disease and three markers on chromosome 16. *J Med Genet* 1987;24:457–61.
8 Lazarou LP, Davies F, Sarfarazi M, Coles GA, Harper PS. Adult polycystic kidney disease and linked RFLPs at the α globin locus: a genetic study in the South Wales population *J Med Genet* 1987;24:466–73.
9 Breuning MH, Reeders ST, Brunner H, *et al*. Improved early diagnosis of adult polycystic kidney disease with flanking DNA markers. *Lancet* 1987;ii:1359–61.
10 Breuning MH, Madan K, Verjaal M, Wijnen JT, Khan M, Pearson P. Human alpha-globin maps to pter-p13.3 in chromosome 16 distal to PGP. *Hum Genet* 1987;76:287–9.
11 Reeders ST, Keith T, Green P, *et al*. Regional localization of the autosomal dominant polycystic kidney disease locus. *Genomics* 1988;3:150–5.
12 Reeders ST, Breuning MH, Ryynanen MA. A study of genetic linkage heterogeneity in adult polycystic kidney disease. *Hum Genet* 1987;76:348–51.
13 Romeo G, Costa G, Catizone L, *et al*. A second genetic locus for autosomal dominant polycystic kidney disease. *Lancet* 1988;ii:7–10.
14 Kimberling WJ, Fain PR, Kenyon JB, Goldgar D, Sujansky E, Gabow PA. Linkage heterogeneity of autosomal dominant polycystic kidney disease. *N Engl J Med* 1988;319:913–18.
15 Wright AF, Carothers AD, Pignatelli PM, *et al*. A study of genetic linkage heterogeneity in 34 adult-onset polycystic kidney disease families. *Hum Genet* (in press).
16 Kunkel LM, Tantravahi U, Eisenhard M, Latt SA. Regional localization on the human X of DNA segments cloned from flow sorted chromosomes. *Nucleic Acids Res* 1982;10:1557–8.
17 Jarman AP, Nicholls RD, Weatherall DJ, Clegg JB, Higgs DR. Molecular characterization of a hypervariable region downstream of the alpha-globin gene cluster. *EMBO J* 1986;5:1857–63.
18 Nakamura Y, Martin C, Krapcho K, *et al*. Isolation and mapping of a polymorphic DNA sequence (pCMM65) on chromosome 16 (D16S84). *Nucleic Acids Res* 1988; 16:3122.
19 Hyland VJ, Suthers GK, Friend K, *et al*. Probe, VK5B, is located in the same interval as the autosomal dominant adult polycystic kidney disease locus, PKD1. *Hum Genet* 1990;84:286–8.
20 Donis-Keller H, Green P, Helms P, *et al*. A genetic linkage map of the human genome. *Cell* 1987;51:319–37.
21 Feinberg AP, Vogelstein B. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 1983;132:6–13.
22 Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 1975;113:237–51.
23 Lathrop GM. Lalouel JM, Julier C, Ott J. Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 1984;81:3443–6.
24 Bear JC, McManamon P, Morgan J, *et al*. Age at clinical onset and at ultrasonographic detection of adult polycystic kidney disease—data for genetic counseling. *Am J Med Genet* 1984;18:45–53.
25 Germino GG, Barton NJ, Lamb J, *et al*. Identification of a locus which shows no genetic recombination with the autosomal dominant polycystic kidney disease gene on chromosome 16. *Am J Hum Genet* 1990;46:925–33.
26 Breuning MH, Snijdewint FGM, Brunner H, *et al*. Map of 16 polymorphic loci on the short arm of chromosome 16 close to the polycystic kidney disease gene (PKD1). *J Med Genet* 1990;27:603–13.
27 Reeders ST. *The search for the PKD1 gene on chromosome 16*. Paper delivered at EEC Concerted Action Workshop on Polycystic Kidney Disease, Paris, 8 June 1990.
28 Gillespie GAJ, Somlo S, Germino GG, Weinstat-Saslow D, Reeders ST. CpG island in the region of an autosomal dominant polycystic kidney disease locus defines the 5′ end of a gene encoding a putative protein channel. *Proc Natl Acad Sci USA* 1991;88:4289–93.

SHORT COMMUNICATION

Evidence for linkage disequilibrium between *D16S94* and the adult onset polycystic kidney disease (PKD1) gene

S E Pound, A D Carothers, P M Pignatelli, A M Macnicol, M L Watson, A F Wright

Genetic mapping of the PKD1 gene in chromosomal region 16p13.3 has depended on the identification of closely linked markers and rare recombinants that suggest the order of loci in relation to meiotic crossover points.[1,2] This has led to the proposal that the PKD1 gene lies in an approximately 750 kb region close to marker loci *D16S84*, *D16S125*, and *D16S94*, and proximal to the hypervariable probe *D16S85* that first detected linkage to *PKD1*.[3,4] However, genetic (meiotic) mapping can sometimes lead to inconsistent localisations, as with the Huntington's disease gene,[5] owing to factors such as rare double crossovers, gene conversion events, gonadal mosaicism, genetic heterogeneity, misdiagnosis, mistyping, and non-paternity. While many of these factors can be minimised or excluded, some are very difficult to exclude and, although rare, become significant compared with the frequency of rare recombination events separating markers from disease genes. In the absence of patients with chromosomal rearrangements that can help to localise the PKD1 gene physically, genetic methods of mapping are critical.

A means of genetically mapping a disease gene, other than by linkage analysis, is by the identification of non-random allelic association (linkage disequilibrium) between a marker and disease locus. This can occur if a mutation causing disease arises in close association with a relatively uncommon marker allele or haplotype, which can then become over-represented in the patient group relative to the general population. The finding of linkage disequilibrium has served to refine the mapping both of the cystic fibrosis gene[6] and of the Huntington's disease gene.[7] However, linkage disequilibrium is expected to occur only if the mutation rate is relatively low or if the frequency of a particular chromosomal haplotype is inflated in the population of affected subjects as a result of previous selective advantage, migration, or population stratification. It requires that many or most affected subjects carry the same original mutation.

It has been suggested that the mutation rate in ADPKD is high,[8] based on the apparently high frequency of unaffected parents of ADPKD subjects and on the high population prevalence. This study was carried out at a time when methods of presymptomatic diagnosis were less sensitive than with ultrasonography and when paternity testing was less readily available. We therefore tested a genetically homogeneous set of 33 ADPKD families for evidence of linkage disequilibrium between the PKD1 locus and closely linked markers *D16S84*, *D16S125*, and *D16S94*.

A set of 34 ADPKD families was ascertained through medical renal units in Edinburgh and Glasgow and the diagnoses established as described.[9–11] Family members were typed for the markers *D16S85* (3'HVR), *D16S84* (CMM65), *D16S125* (26-6PROX), and *D16S94* (VK5). Multipoint heterogeneity analysis was carried out to determine the proportion of families unlinked to the PKD1 region of chromosome 16.[10] The sample was found to show evidence of linkage heterogeneity but 90 to 95% of the kindreds were linked to *PKD1*. A single unequivocally unlinked family was removed from the analysis, leaving 33 families for genetic mapping of the PKD1 locus.[11] In 28 families the haplotypes could be determined unambiguously, in one family the haplotype could not be established, and in the remaining four this was done by minimising the number of crossovers. The normal chromosomes used to establish control allele frequencies were derived from the unaffected spouses of family members.

The results are shown in the table. No evidence of allelic association was found between *PKD1* and either *D16S84* or *D16S125*, both of which show close genetic linkage to *PKD1*.[1,2,11] In the case of *D16S94*, there is evidence of linkage disequilibrium, with the affected haplotypes in seven kindreds showing the 1·6 kb allele (allele 1) and in 22 kindreds the 1·3 kb allele (allele 2) compared with the control population, who showed 38 1·6 kb alleles and 34 1·3 kb alleles. This result is significant at the 2% level (Yates's corrected $\chi^2 = 5{\cdot}75$). The allele frequencies in the control population are in good agreement with those published previously.[2]

The suggestion of linkage disequilibrium between *PKD1* and *D16S94* has implications for the genetic localisation of *PKD1*. Any initial disequilibrium occurring when a mutation arises is dissipated at the rate of $(1-r)^n$ in n generations, where r is the recombination

MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU.
S E Pound
A D Carothers
P M Pignatelli
A F Wright

Department of Medicine, Royal Infirmary of Edinburgh, Lauriston Place, Edinburgh.
A M Macnicol
M L Watson

Correspondence to Dr Wright.

Received 8 October 1991.
Accepted 22 October 1991.

Frequency of marker alleles in haplotypes of ADPKD (PKD1+ chromosomes compared with the general population. Yule's association coefficient provides a measure of the degree of disequilibrium.[6] χ^2 values were derived using Yates's correction.

	D16S84		*D16S125*		*D16S94*	
	Affected	Control	Affected	Control	Affected	Control
Allele 1	13	28	9	13	7	38
Allele 2	15	38	15	49	22	34
χ^2	0·020		1·69		5·75	
A	0·081 ± 0·225		0·387 ± 0 223		0·557 ± 0·170	
p	NS		NS		<0·02	

fraction separating marker and disease locus. If r is very small, the rate of approach to equilibrium is slow. Similarly, a recent mutation will take a finite time to reach equilibrium with adjacent markers, even when r is significant. In practice, a plateau of linkage disequilibrium can define an area of up to several hundred kilobases of DNA on each side of the gene of interest, in the absence of recombinational hot spots. The identification of meiotic crossover points by haplotype analysis has led to the proposal that *PKD1* lies between *D16S84* and *D16S125*, a region shown by pulsed field electrophoresis to contain less than 750 kb of DNA and found to contain a large number of genes.[4] The more proximally located marker *D16S94* lies just outside this area. If the presence of linkage disequilibrium can be confirmed by other groups with ethnically homogeneous patient populations, it would argue for a detailed physical mapping effort in the proximal part of the *D16S84-D16S125* interval, although further marker studies are required to establish the extent of the region of disequilibrium. The failure to find any disequilibrium with *D16S125*, located about 50 to 100 kb distally, could be because a common PKD1 mutation arose in association with the more frequent *D16S125* allele, making it more difficult to detect disequilibrium, or because the *Taq*I polymorphism is itself subject to significant mutational change.

The high population prevalence of ADPKD (1 in 1000), 90 to 95% of which are the result of PKD1 gene mutations,[4 10] is consistent with a high mutation rate but since the reproductive fitness of ADKPD patients is also high, this is not necessarily so. Additional factors, such as reproductive compensation, previous selective advantage, and genetic drift, may also contribute to the high prevalence. Linkage disequilibrium is not expected to occur if the mutation rate is high since each new mutation is likely to occur on a different chromosomal background. The results described in these ADPKD families therefore add support to the idea that the high prevalence is maintained by a high reproductive fitness or other factors referred to above, in the presence of an intermediate or low mutation rate. These issues await further clarification when the gene is isolated.

The financial support of The Wellcome Trust is gratefully acknowledged. We would also like to thank Drs J D Briggs and B Junor for access to their patients. SEP is in receipt of an MRC Training Fellowship.

1 Germino GG, Barton NJ, Lamb J, *et al.* Identification of a locus which shows no genetic recombination with the autosomal dominant polycystic kidney disease gene on chromosome 16. *Am J Hum Genet* 1990;46:925–33.
2 Breuning MH, Snijdewint FGM, Brunner H, *et al.* Map of 16 polymorphic loci on the short arm of chromosome 16 close to the polycystic kidney disease gene (PKD1). *J Med Genet* 1990;27:603–13.
3 Reeders ST, Breuning MH, Davies KE, *et al.* A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. *Nature* 1985;317:542–4.
4 Gillespie GAJ, Somlo S, Germino GG, *et al.* CpG island in the region of an autosomal dominant polycystic kidney disease locus defines the 5' end of a gene encoding a putative proton channel. *Proc Natl Acad Sci USA* 1991;88:4289–93.
5 MacDonald ME, Haines JL, Zimmer M, *et al.* Recombination events suggest potential sites for the Huntington's disease gene. *Neuron* 1989;3:183–90.
6 Kerem B, Rommens JM, Buchanan JA, *et al.* Identification of the cystic fibrosis gene: genetic analysis. *Science* 1989;245:1073–80.
7 Snell RG, Lazarou LP, Youngman S, *et al.* Linkage disequilibrium in Huntington's disease: an improved localisation for the gene. *J Med Genet* 1989;26:673–5.
8 Dalgaard OZ. Bilateral polycystic disease of the kidneys: a follow-up of two hundred and eighty four patients and their families. *Acta Med Scand* 1957;328:1–255.
9 Reeders ST, Breuning MH, Ryynanen MA, *et al.* A study of genetic linkage heterogeneity in adult polycystic kidney disease. *Hum Genet* 1987;76:348–51.
10 Wright AF, Carothers AD, Pignatelli PM, *et al.* A study of genetic linkage heterogeneity in 34 adult-onset polycystic kidney disease families. *Hum Genet* (submitted).
11 Pignatelli PM, Pound SE, Macnicol AM, *et al.* Multipoint mapping of adult onset polycystic kidney disease (PKD1) on chromosome 16. *J Med Genet* (in press).