NARROWING OF THE AUTOSOMAL RECESSIVE POLYCYSTIC KIDNEY DISEASE CRITICAL REGION IN A NEWFOUNDLAND FAMILY

TOBY FROST

NARROWING OF THE AUTOSOMAL RECESSIVE POLYCYSTIC KIDNEY DISEASE CRITICAL REGION IN A NEWFOUNDLAND FAMILY

by

Toby Frost

A thesis submitted to the School of Graduate Studies in partial fulfilment of the requirements for the degree of Master of Science

Department of Biochemistry

Memorial University of Newfoundland

January 2000

St. John's Newfoundland

ABSTRACT

A Newfoundland family, consisting of three nuclear families, was identified as having polycystic kidney and liver disease and found to be descended from two founding ancestral pairs. Using homozygosity mapping with polymorphic microsatellite markers, three disease loci, NPH-1, -2 and mouse bpk/jcpk were excluded and linked the disease in sibships A and B to the ARPKD disease loci on chromosome 6. Due to a meiotic recombination in one of the individuals in sibship A, the ARPKD critical region was narrowed to a 1cM region. The disease in sibship C was not linked to either of the abovementioned loci and kidney ultrasound suggested that the disease was not ADPKD. Because the gene which causes ARPKD is, as yet, unknown, it was necessary to attempt to identify it through positional cloning. A PAC contiguous map was constructed to allow the examination of smaller, individual pieces of DNA within the region. Each clone was subcloned and screened for microsatellites. Three were identified however none was variable in this family. The complete sequence of clone 108c2 had been published to the public domain and examination revealed a large repeat region which was analyzed and found to be polymorphic. This marker allowed the ARPKD region to be further refined to an area of approximately 560kb in size, completely covered by PAC clones. This region contains 10 known gene-oriented clusters representing different transcripts including NFYA, TFAP2B, MDFI and APOBEC2 (http://www.ncbi.nlm.nih.gov/ UniGene).

ACKNOWLEDGMENTS

I would like to sincerely like to thank the faculty and staff of the Memorial University Department of Biochemistry who were always there when I needed a hand, information or advice. I would also like to thank those people with whom I shared lab space every day for two years; especially Sylvia Bartlett who provided a mountain of technical advise and Terry-Lynn Young who was there for me from start to finish helping with lab work and thesis writing. In addition, I would like to acknowledge my fellow graduate students, we created a close-knit social group, we were able to give each other advice and we had some great times. Thanks for all the laughs! To my supervisory committee, Dr. Jane Green and Dr. Patrick Parfrey, I appreciate all the work you put in providing me with technical advise and thesis corrections. Finally, I would like to thank my supervisor, Dr. William S. Davidson.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Adenosine	**A**
Adenosine Triphosphate	**ATP**
Alpha	**α**
Ammonium Chloride	**NH_4Cl**
Autosomal Dominant Polycystic Kidney Disease	**ADPKD**
Autosomal Recessive Polycystic Kidney Disease	**ARPKD**
Bacterial Artificial Chromosome	**BAC**
Base Pair	**bp**
Carbenicillin	**Carb**
complementary DNA	**cDNA**
CentaMorgan	**cM**
Complementary DNA	**cDNA**
Contiguous Map	**Contig**
Curie	**Ci**
Cytosine	**C**
Degrees Celsius	**°C**
Deoxy/Dideoxy Nucleotide Triphosphates	**d/ddNTP**
Distilled Water	**dH_2O**
Deoxyribose Nucleic Acid	**DNA**

End Stage Renal Disease	**ESRD**
Epidermal Growth Factor Receptor	**EFG-R**
Ethylenediaminetetra-acetic Acid Disodium Salt	**EDTA**
Gamma	**γ**
Grams	**g**
Guanosine	**G**
Homozygosity by Descent	**HBD**
Hydrochloric Acid	**HCl**
Identical by Descent	**IBD**
Isopropyl-β-thiogalactopyranoside	**IPTG**
Juvenile Nephronophthisis	**NPH**
Kanamycin	**Kan**
KiloBase pair	**kb**
Lambda	**λ**
Liter	**L**
Log of the Odds	**LOD**
MegaBase pair	**Mb**
Messenger Ribonucleic Acid	**mRNA**
MicroLiter	**μL**
MilliCurie	**mCi**
MilliLiter	**mL**

MilliMeter	**mm**
MilliMolar	**mM**
Minutes	**min.**
Molar	**M**
Oligonucleotide	**Oligo**
Personal Identification	**PID**
Phosphorus-32	**^{32}P**
Phage Artificial Chromosome	**PAC**
PicoMolar	**pM**
Polymerase Chain Reaction	**PCR**
Polynucleotide Kinase	**PNK**
Potassium Acetate	**KAc**
Revolutions per Minute	**rpm**
Seconds	**Sec.**
Sequence Tagged Site	**STS**
Shrimp Alkaline Phosphatase	**SAP**
Sodium Chloride	**NaCl**
Sodium Phosphate/Sodium Chloride/EDTA Buffer	**SSPE**
Sodium Citrate/Sodium Chloride Buffer	**SSC**
Sodium Dodecyl (Laural) Sulfate	**SDS**
Sodium Hydroxide	**NaOH**

Sodium Phosphate	**NaH_2PO_4**
Theta (Recombination Fraction)	**θ**
Thin Layer Chromatography	**TLC**
Thymidine	**T**
Times Force of Gravity	**x g**
Tris/Boric Acid/EDTA Buffer	**TBE**
Tris/EDTA Buffer	**TE**
Tris/EDTA/NaCl Buffer	**TNE**
Trizma Base	**Tris**
Transforming Growth Factor	**TGF**
Ultra Violet	**UV**
Units	**u**
Family Unknown 1	**U1**
Watt	**w**
Yeast Artificial Chromosome	**YAC**

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

1. INTRODUCTION

1.1 GENETICS AND DISEASE

Genetic diseases are caused by a change in the sequence of DNA, differential methylation of a DNA strand, or through chromosomal rearrangements. A change in sequence within a gene may result in a change in the transcribed RNA and may affect the resulting protein sequence. Large insertions and deletions will significantly alter a gene product which can affect its function by deleting functional domains or altering the structure of the product. Deletions may also affect proper cellular localization, proper splicing or the rate of transcription if the deletion involves a promoter region. If the gene product performed an essential function then the mutation may cause systemic manifestations. Single base pair insertions or deletions within an exon will cause a change in the reading frame and cause the translated product be altered or truncated due to the creation of a "stop" codon, resulting in a shortened gene product. Single base pair substitutions, which do not alter the reading frame of the DNA, may cause an altered amino acid sequence which can change the function of the protein. A mutation in a noncoding region of the genome may also cause disease if the mutation occurs in an area of the genome corresponding to a regulatory region (Strachan and Read, 1999).

Different diseases are described as having either a dominant or recessive mode of inheritance. Dominant inheritance occurs when the disease is expressed when only one copy

of the pair of genes is mutated. In these diseases the mutated gene product will often acquire an abnormal function, explaining why only one copy of the gene is required to cause the disease. In recessive inheritance, on the other hand, the disease is expressed only when both copies of a pair of genes are mutated. In these cases the gene product has often lost an essential function, thus two mutated genes are required to fully express the disease phenotype. Usually when one copy of a mutated gene is inherited the other copy is able to compensate for the loss through increased expression, often through a feedback loop system. By identifying a gene and the nature of the mutation which causes a disease, it is possible to predict the mutated gene product and compare it to the normal product. Once a gene has been cloned and the function determined, it may then be possible to develop a medical treatment in order to prevent the disease or alleviate the symptoms.

The type of disease can often be determined by simply examining the pedigree of affected individuals. The five important Mendelian pedigree patterns are autosomal dominant, autosomal recessive, X-linked dominant, X-linked recessive and Y-linked; each of which is subject to a variety of complications. An individual with an autosomal dominant disease usually has at least one affected parent and approximately 50% of the children of an affected individual will inherit the disease, and affected individuals are seen throughout the pedigree in several generations. For autosomal recessive diseases, affected individuals usually have unaffected parents and appear in a single generation of the pedigree. This is especially true for marriages in which there is some degree of consanguinity. X-linked recessive diseases affect mostly males as males are hemizygous for the X chromosome and

they usually have unaffected parents. Females are affected only if the father is affected and the mother is a carrier, but there is never any father to son transmission of the disease. X-linked dominant diseases affect more females than males and half of the offspring of an affected female will also inherit the disease as there is a 50% chance that she will pass the mutated chromosome to her offspring. In the case of an affected male, all of his daughters will be affected but none of his sons. This is because the X-chromosome is inherited by his daughters and only the Y-chromosome is inherited bt his sons. Finally, Y-linked diseases affect only males. All affected males have an affected father, provided the mutation is not new (Strachan and Read, 1999). These general patterns can be complicated by such phenomena as late onset diseases where a disease does not manifest until late in life or decreased penetrance where not all individuals who inherit a mutation develop a phenotype.

Genetic research is presently going through a revolution due to the rapid expansion of sequence data available from public databases. This information allows researchers to compare their data to other data sets quickly and easily. It is now possible to take nucleotide sequence data of interest and search for similar sequences at the nucleotide or protein level, allowing for the discovery of functional domains or expressed sequences. Homology searches of genes discovered in simpler organisms have lead to the discovery of many new human genes. Expressed sequence tag and sequence tagged site databases allow for the identification of transcribed sequences within a region. (Pandey and Lewitter, 1999).

1.2 FINDING A DISEASE GENE

1.2.1 Positional Cloning

Over the past 20 years many disease genes have been identified by using molecular cloning techniques, approximately 1677 according to the Online Mendelian Inheritance in Man (http://www3.ncbi.nlm.nih.gov/Omim/). "Traditional" methods required that some information on the gene function was known. For example the gene for phenylketonuria was discovered because of prior knowledge of the biochemical deficiency (Robson et al., 1992). However, for the majority of the thousands of genetic diseases identified, there is no functional information available on the nature of the gene product. The technique of positional cloning allows for the identification of a disease gene by first identifying a map position (Collins, 1992a).

The initial step in positional cloning is the localization of the gene in question to a relatively small area of the genome. This involves screening large multiplex families in which there is a well defined phenotype. Polymorphic markers are used to screen the DNA of the individuals of the family in order to determine the inheritance of each area of the genome for affected and unaffected individuals in relation to the trait of interest (Leppert, 1990; Collins, 1992b). The polymorphic markers used are usually simple di- tri- or tetra-nucleotide repeats whose locations in the genome are known. These markers, also known as microsatellite markers, are generally found every 100kb throughout the human genome (Dib et al., 1996; The Utah Marker Development Group, 1995). The length of the repeat is

quite variable yet relatively stable within families. Because they are made up of a repeating sequence, stem loops may form during replication and cause the newly synthesized DNA to be either longer or shorter than the original repeat sequence. The markers have unique sequences on either end of the repeat, allowing for the creation of oligonucleotide primers which can be used to amplify and study specific markers.

The only drawback to this method is the size of the human genome which has a haploid size of three billion nucleotides. This means that 98% to 99% of the markers tested will not be linked to the gene (Collins, 1992b). To date, there have been 171 disease genes discovered using the method of positional cloning (OMIM, http://www3.ncbi.nlm.nih.gov/Omim/)

1.2.2 Homozygosity Mapping

During the meiosis I phase of gamete production, following replication, members of each of the chromosomal pairs, maternal and paternal, come together and align at the metaphase plate of the cell. During anaphase I, the bivalent chromosomes separate and migrate to opposite poles of the cell prior to division. For each of the 23 pairs of chromosomes, which homolog ends up in each daughter cell is random, creating 8.4×10^6 possible combinations (Strachan and Read, 1999). This satisfies Mendel's law of independent assortment. However Mendel was unable to predict the phenomenon of genetic recombination which occurs during gametogenesis. During prophase of meiosis I, the paired homologous chromatids randomly exchange genetic material. This exchange involves a

physical break in each of the homologous pair of chromosomes and rejoining to the sister chromosome. Thus, the paternally derived chromosome will exchange genetic material with the maternally derived chromosome. As this process of recombination occurs at ensuing generations, it is only the most closely linked chromosomal regions which will remain linked. Identifying these regions is the basis for gene mapping and allows geneticists to refine an area of the genome which contains a disease gene.

Individuals with a rare recessive disease in consanguineous families are likely to be homozygous for the disease causing mutation as well as for an ancestral haplotype (a common set of alleles for markers linked to the disease locus) because of inheritance from a common ancestor. This is known as homozygosity by descent (HBD) (Lander and Botstein, 1987). The more remote the common ancestor, the smaller the shared portion of the genome between relatives and the greater the significance for linkage to a disease gene if HBD can be identified (Strachan and Reid, 1999). Homozygosity mapping involves finding areas of the genome which are HBD in affected individuals but not unaffected individuals (Strachan and Read, 1997).

Markers are initially used to screen segments of the genome identified as candidate loci, segments containing potential disease inducing genes. When evidence for homozygosity among all affected individuals is found for one or more of the polymorphic markers, the data are analyzed by computer to determine the probability that the markers are linked to the disease being investigated. This is calculated by determining the probability that homozygosity of the markers occur in the affected individuals by chance assuming a

particular recombination frequency (θ). The calculation is then redone assuming that the marker and gene of interest are linked and the ratio of the two probabilities (linked and unlinked at a particular θ) is determined. This ratio, expressed in Log_{10}, depicts the ratio of odds for and against that degree of linkage. This is known as a LOD (Log of the odds) score and it is expressed as z (Morton, 1955). If z is greater than or equal to 3 it is considered evidence for linkage and represents a minimum of 1000:1 odds that the loci are linked. A negative LOD score indicates a recombination has occurred between a marker and gene of interest while LOD score ≤ -2 is considered conclusive evidence against linkage between a marker and gene (Thompson et al., 1991 and Watson et al., 1992). Once linkage with a set of markers has been determined, the area containing the disease gene, known as the critical region and defined by flanking markers, is usually quite large. The next step in the process is to narrow the region down as much as possible. To accomplish this, the region is further examined for the presence of microsatellite repeats until the area is saturated with a large number of polymorphic markers. These markers are then used to identify meiotic recombination events which would define the boundaries of a refined critical region containing the mutated gene. The critical region can now be defined by the smallest area known to be homozygous for all affected individuals and none of the unaffected individuals.

When the critical region has been sufficiently refined, the next step is to identify genes within this region which could cause the disease. The Human Genome Project, whose goal it is to completely map the human genome, can provide useful information on which genes are in a specific area. However, this project is far from over; not all genes are mapped

and the function of other genes are unknown. A working draft is expected to be completed sometime this year. However, the complete reference sequence is not expected until 2003(Francis Collins, Director of NIH, in a public address to the 49th annual meeting of "The American Society of Human Genetics", San Francisco, USA). Therefore, it may not be possible to find the gene of interest in any database. On average, a 1 Mb region of human DNA contains approximately 20 to 30 genes which need to be isolated and tested for the presence of mutations (Fields et al., 1994). This process of identifying a gene by location first and then discovering its function, as opposed to beginning with a known gene product and using the amino acid sequence to isolate the gene, is known as positional cloning. The first gene to be successfully linked to a polymorphic marker with a known locus in the human genome in this manner was the cystic fibrosis gene (Tsui et al., 1985)

1.2.3 Physical Mapping

The goal of physical mapping is to refine a genes's location to a small enough region, usually 1cM, so that all expressed sequences can be identified and tested to see which is the gene of interest. To further refine a gene's location, the critical region must first be cloned using a phage (PAC), bacterial (BAC), cosmid, P1, or yeast artificial chromosome (YAC) (Neuhausen et al., 1994; Shizuya et al. 1992; Feingold et al., 1990; Carrano et al., 1989). Cosmids are easy to work with; however, the DNA insert which they are able to accommodate is quite small, only about 40 to 45Kb (Whittaker et al., 1988). In contrast, YACs are able to accommodate relatively large inserts, consisting of up to 1 Mb. However,

they are difficult to work with as they are difficult to grow and a large number of them undergo recombination (Burke et al., 1987; Traver et al., 1989). Presently, PACs and P1s are between YACs and cosmids when considering advantages and disadvantages, as they are quite easy to work with and are able to handle an insert of approximately 120 Kb (Sternberg et al., 1992; Ioannou et al., 1994) (Fig. 1). The DNA clones are then screened for the presence of genetic markers which are known to map to the critical region. A clone which contains a marker that maps to the critical region is isolated and tested with other markers in order to determine how much of the critical region the particular clone covers and if it overlaps with other clones. Overlapping clones contribute to the construction of a physical map referred to as a contiguous map, or contig.

Once the contig has been produced, a search for individual genes within the area may begin, involving the use of one or more techniques. One technique commonly used is the screening of complementary DNA (cDNA) libraries. cDNA libraries are created from the mRNA of specific tissue types at a specific age, fetal kidney for example. Expressed sequences from a specific area can be isolated by using radiolabeled PAC or YAC DNA as a probe (Lovett et al., 1991). A more involved technique consists of sequencing the whole area and performing a gene recognition search of genomic databases (Uberbacher and Mural, 1991). Exon trapping is a technique whereby genomic fragments are cloned into a plasmid where the cloning site lies between a splice donor and a splice acceptor flanked by two known sequences. The plasmid is then transfected into a suitable eukaryotic cell line where the cellular machinery transcribes the insert DNA along with the known flanking sequences.

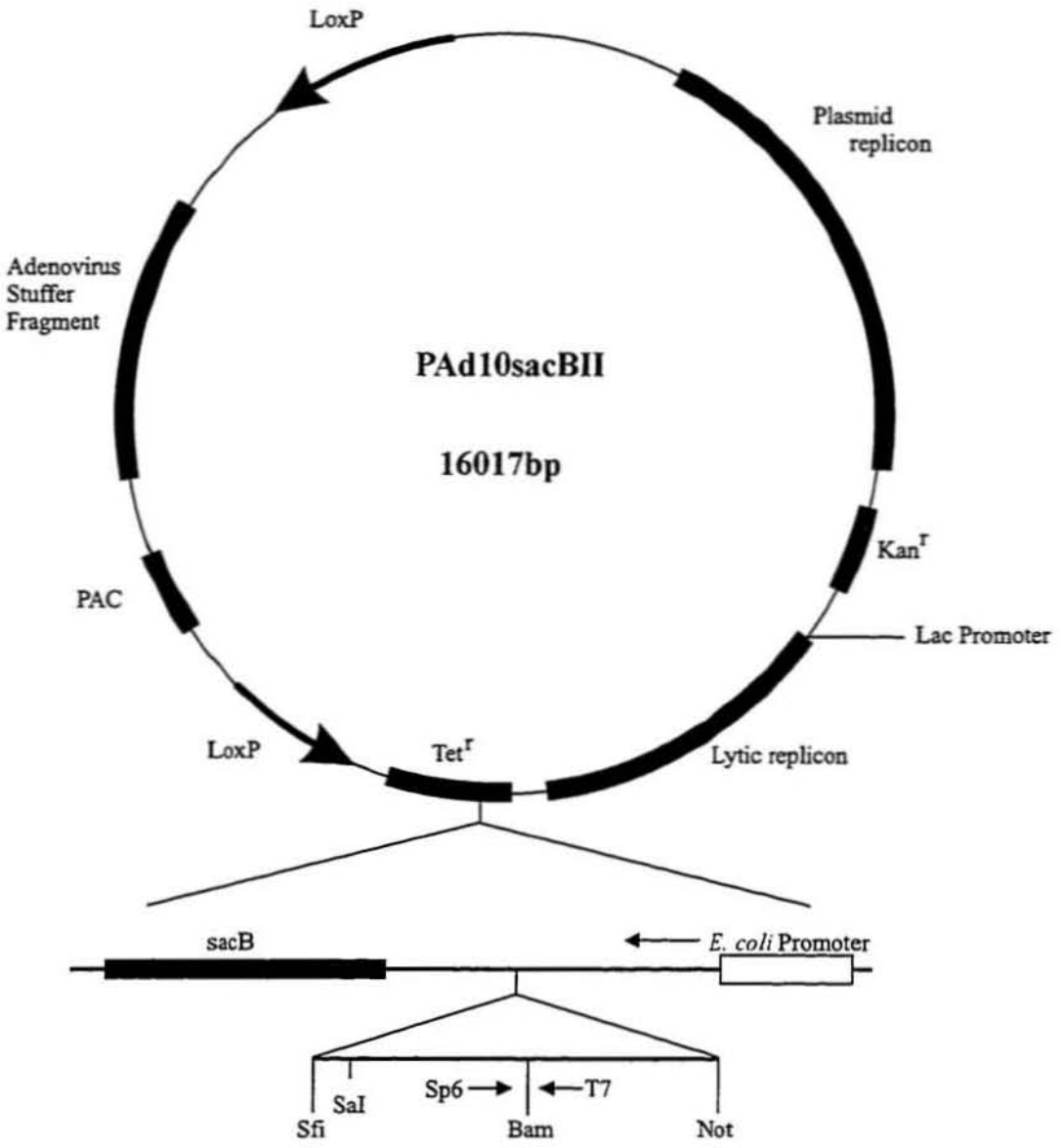


Figure 1: An illustration of the pAd10sacBII PAC vector depicting its major features. Insert DNA is cloned into a BamHI site by SP6 and T7 sequences which facilitate sequencing the ends of the insert. The plasmid also contains a lytic replicon which is under the control of the Lac promoter. This enables the induction of plasmid synthesis resulting in up to 20 copies per cell.

Once spliced, the unknown exon is trapped between two known sequences where is can be used as a template to make cDNA. The trapped exon may then be sequenced and analyzed or it may be used as a probe to search for full length cDNAs (Duyk et al., 1990). Finally, the candidate gene approach involves determining which genes have already been mapped to the area of interest, researching their function, and determining if one of them could be the cause of the disease being studied. This latter approach is important as there have been very few human disease genes which have been isolated solely on the basis of isolation from a particular area and identification of mutations (Papadopoulos, 1995).

The final step in the identification of a disease gene after candidate genes have been identified is the search for a mutation within one of these genes. The occurrence of this mutation in all the affected individuals and none of the unaffected individuals is evidence that the particular gene causes the disease. However, before a gene is considered to be the true cause of a disease it must first be tested in other families with the same disease. This is done in case the mutation in the candidate was a coincidence and not the actual cause of the disease. Once this has been determined it is then necessary to test other families with the same disease for similar or novel mutations in the gene. The most accurate method of finding genetic mutations is direct sequencing. However, there are other techniques such as denaturing gels or single stranded conformational polymorphism analysis which can give consistently reliable results (Orita et al., 1989). Once the gene is identified, a search for the exact function of the gene product begins in hope that a way to supplement its function will be available (Papadopoulos, 1995).

1.3 CLINICAL PRESENTATION OF THE FAMILY

1.3.1 Geneology

A large multiplex Newfoundland family of 153 individuals (U1) was identified with six children (PID 49, 50, 51, 53, 54 and 55) in three interrelated sibships A, B and C affected with a renal cystic-biliary disease complex (Fig. 2). The parents of the affected individuals in sibship A are 3rd cousins, sharing their maternal great-great-grandparents (PID152 and 153). In sibship C the parents of the affected individual (PID 55) are related as 2nd cousins, sharing paternal great-grandparents (PID 150 and 151). The parents of individuals in sibship B are related to both sibships A and C. The father of sibship C(PID 40) is the 2nd cousin to the father of sibsip A (PID 29) and the mother of sibship B (PID 41) is the 1st cousin to the father of sibship C(PID 38). Clinical data were collected by Dr. Patrick Parfrey, Patient Research Centre, Health Sciences Centre, St. John's, Canada, and Dr. Benvon Cramer, Janeway Child Health Centre, St. John's, Canada. The clinical data were reviewed by Dr. Lisa Guay-Woodford, University of Alabama, Birmingham, USA.

1.3.2 Sibship A

The proband (PID 49), a male, presented at 2 years of age with failure to thrive and hepatosplenomegaly. Liver ultrasound showed an abnormal parenchyma and cysts. At age 5 he had portal hypertension and varicies in addition to congenital hepatic fibrosis identified through a liver biopsy. Kidney ultrasound at age 8 indicated a non-functioning left kidney.

The right kidney was hypertropied with cystic changes and a mild increase in echogenicity of the right kidney with the preservation of corticomedullary differentiation. This individual required a splenorenal shunt at age 13 and started dialysis soon after. A kidney transplant was performed at age 14 which is still functioning at age 24. Individual PID50, female, presented with an enlarged liver and spleen at age 1. Liver ultrasound showed increased echogenicity and an abnormal parenchyma. A liver biopsy, performed at age 8, revealed extensive portal tract fibrosis with bile duct proliferation indicating congenital hepatic fibrosis. Echogenicity in left kidney increased in cortex but the right kidney remained normal. A kidney ultrasound, age 16, revealed normal corticomedullary differentiation, however she did required a splenorenal shunt. Serum creatinine was 53 μmol/L at age 17. Individual PID 51, female, presented with an enlarged liver at age 2. At age 5, liver ultrasound revealed cysts in right lobe with diffuse increase in size and increased echogenicity and enlarged spleen with umbilical vein varices. Kidney ultrasound indicated a mild increase in cortical echogenicity with normal corticomedullary differentiation without cysts. At age 10, serum creatinine level was 57 μmol/L.

1.3.3 Sibship B

The affected female (PID 53) presented at age 6 with symptoms of hepatosplenomegaly and congenital hepatic fibrosis was diagnosed in a liver biopsy. Liver ultrasound showed increased echogenicity while renal ultrasound revealed large, speckled, echodense kidneys with a loss of corticomedullary differentiation but no discrete cysts. At

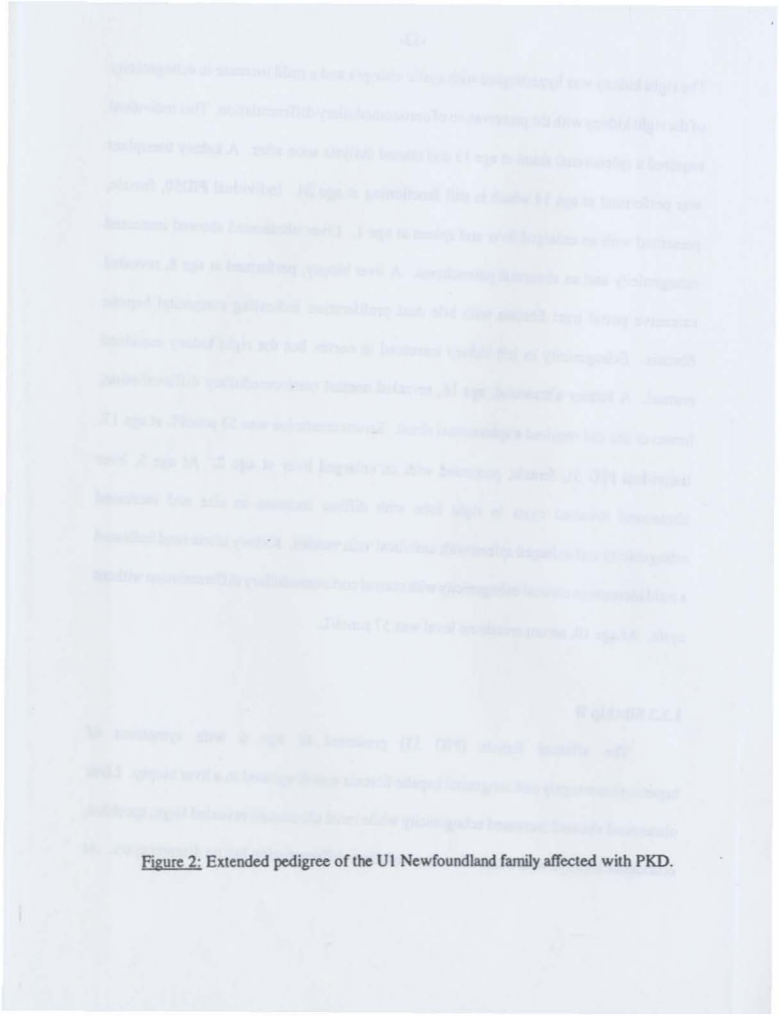


Figure 2: Extended pedigree of the U1 Newfoundland family affected with PKD.

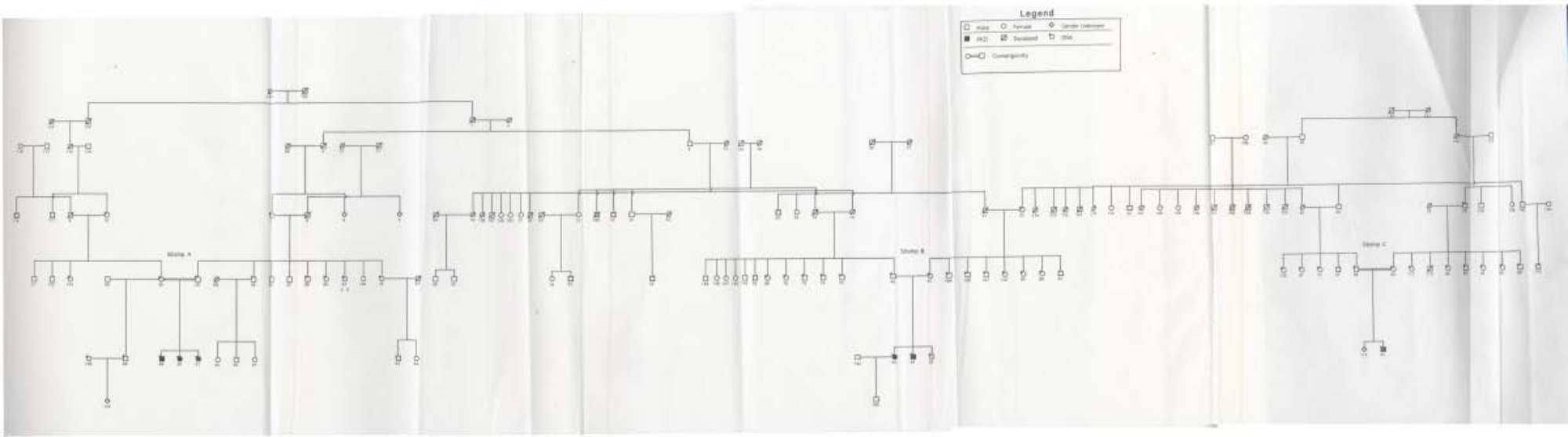

Legend

Figure 2 Pedigree.

age 16, serum creatinine level was 125 μmol/L. At age 22, she had asymptomatic portal hypertension and serum creatinine was 217 μmol/L, indicating decreased kidney function. The affected male (PID 54) presented with hepatosplenomegaly at age 8. Liver ultrasound showed increased echogenicity with no discrete cysts. Biopsy revealed bile duct proliferation with minor fibrosis. Renal ultrasonography displayed large speckled echodense kidneys with a loss of corticomedullary differentiation, and no cysts. At the age of 15 years he had a splenorenal shunt and a year later he had a kidney transplant which continues to function at age 21.

1.3.4 Sibship C

The single affected individual in sibship C (PID 55) had hernia repair at age 2 and abdominal ultrasound. Multiple large cortical and medullary cysts were observed on his kidneys, consistent with ADPKD, with the preservation of the renal architecture. Liver ultrasound was normal. At age 7, serum creatinine was 68 μmol/L. As of yet this individual has not displayed any liver or spleen dysfunction (Table 1).

1.3.5 Suitability for genetic study

This family is ideal for a genetic study due to several factors relating to the pedigree structure. First, the pedigree can be traced back six generations giving a good record of the family history. Second, the family is fairly large and has several affected individuals in three different sibships allowing for comparison of several affected individuals. Third, within the

Table 1: Summary of manifestations observed in the affected individuals in each of the three sibships of the U1 family.

Sibship	Kidney involvement	Liver involvement	Spleen involvement	Esophageal varicies
A	✔	✔	✔	✔
B	✔	✔	✔	
C	✔			

family as a whole, there are two consanguineous marriages making it very likely that the affected individuals are HBD in the disease area (Lander and Botstein, 1987; Sheffield et al., 1998).

1.4 Possible Disorders

1.4.1 Disease Considerations

The family pedigree did reveal useful information when considering possible diseases. It is clear from Fig. 1 that there is no bias towards males or females and is therefore not linked to the X chromosome. There is no evidence that maternal or paternal inheritance plays a role which rules out the involvement of mitochondrial DNA or the Y chromosome. The observation that the disease only appears in one generation and there are two consanguineous marriages led to the conclusion that it is of an autosomal recessive nature. The pathological features included enlarged spleen, enlarged polycystic kidneys with a lack of corticmedullary differentiation, enlarged cystic liver and high serum creatinine levels all of which occur within the first few years of life. Two human polycystic kidney diseases match this profile; autosomal recessive polycystic kidney disease (ARPKD) and juvenile nephronophthisis (NPH) both of which are known to be caused by autosomal genes and are recessive in nature. Another human disease which closely follows the list of phenotypic features found here is autosomal dominant polycystic kidney disease (ADPKD). While ADPKD is not a recessive disease, the features closely match the list of phenotypes found

in the U1 family. Finally, research in polycystic kidney disease has focused on similar diseases in mice in an attempt to find human homologs. The mouse mutations *bpk* and *jcpk* have been found to cause polycystic kidneys with different severities and have recently been found to be allelic (Guay-Woodford et al., 1996). This finding was quite intriguing as the mutations are found in an area of the mouse genome known to be syntenic to part of human chromosome 10.

1.4.2 Juvenile Nephronophthisis

Juvenile Nephronophthisis (NPH) (MIM # 256100) was first reported in the literature in 1945 by Smith and Graham; however, the first description is attributed to Fanconi et al. in 1952 (Gusmano et al., 1998). NPH is characterized by its recessive inheritance and the formation of cysts at the corticomedullary junction, chronic tubulointerstitial nephritis. It was later discovered that NPH is also associated with congenital hepatic fibrosis (Proesmans et al., 1975; Witzleben and Sharp, 1982). In most cases, children initially present at around age 4 years with polyurea, polydipsia and isosthenuria due to a reduction in the ability of the kidney to concentrate urine (Konrad et al., 1998; Gusmano et al., 1998). NPH symptoms include: slowly progressing renal failure with shrunken kidneys and diffuse interstitial lesions, of 1-15mm in diameter, eventually leading to end stage renal disease (ESRD) at median age of 13 years (Haider et al., 1998; Fick and Gabow, 1994). Ultrasonography reveals medullary cysts and a loss of corticomedullary differentiation in kidneys. Histological analysis reveals a thickening in the basement membranes. Analysis using anti-

basement membrane antibodies against proteins such as laminin and type IV collagen antibody reveals either normal or a complete lack of binding (Konrad et al., 1998; Gusmano et al., 1998). NPH is estimated to be the cause of 10 - 14% of all renal failure in children (Gusmano et al., 1998). Although purely renal, NPH can be associated with nonrenal diseases as in Senior-Løken disorder (MIM # 266900), nephronophthisis and retinal dystrophy. Bone abnormalities have also been reported to be associated with NPH, some of these include: cone shaped epiphyses, and iliac and acetabular malformations (Gusmano et al., 1998; Konrad et al., 1998).

In 1993 one of the genes causing NPH (NPH-1) was localized to a region on chromosome 2q12-q13 (Antignac et al., 1993). NPH-1 has since been cloned and identified as a novel gene (Hildebrandt, 1998). In 1998 the NPH-2 locus was discovered through homozygosity mapping using a large multiplex Bedouin family whose affected members displayed the NPH phenotype but did not map to the NPH-1 locus on chromosome 2 (Haider et al., 19987). The NPH-2 locus maps to a 12.9 cM region on chromosome 9q22-31, flanked by the markers D9S280 and GGAT3G09.

1.4.3 Autosomal Dominant Polycystic Kidney Disease

ADPKD (MIM # 600666) is present throughout the world and its prevalence is estimated to be approximately 1:200-1:1000 (Bycroft et al., 1999; Calvet, 1998; Merta et al., 1997). There are at least three known loci responsible for ADPKD on chromosomes 16p 13.3 (known as the PKD1 locus), chromosome 4q13-q23 (PKD2 locus) and at least one

other, as yet unknown, locus. The disease can be quite variable depending on which gene is involved, resulting in ESRD at significantly different times in life. However, even individuals in the same family can display remarkable variability in the severity or age of onset of the disease, including presentation as congenital polycystic kidney disease and hepatic fibrosis (Chauveau et al., 1997). Novel mutations of ADPKD have been identified (Kim et al., 2000; Iglesias et al., 2000).

1.4.4 Mouse bpk/jcpk mutations

The mutations *bpk/jcpk* are actually genetic mutations which occur in the mouse and give a phenotype which mimics several human PKD mutations including autosomal recessive polycystic kidney disease (ARPKD). The *bpk* mutation is a less severe form of the disease and arose spontaneously in an inbred population of laboratory mice and thus most likely represents a point mutation, whereas the *jcpk* mutation was induced by the mutagen chlorambucil which is known to cause deletions and rearrangements in genetic material (Iakoubova et al., 1994; Bryda et al., 1996; Moyer et al., 1994). It was found that these two mutations, which produce different phenotypes, occur within the same discrete interval of the mouse genome and it is now thought that they are allelic (Guay-Woodford et al., 1996). The locus of the mouse genome where this occurs is syntenic to a 17 cM region on human chromosome 10. The notion that differences in the severity of ARPKD could be due to allelic differences is intriguing.

1.4.5 Autosomal Recessive Polycystic Kidney Disease

ARPKD (MIM # 263200) is one of the most common hereditary renal cystic diseases in children, involving the kidneys and the biliary tract. As many as 50% of children born with ARPKD die within hours of birth due to respiratory failure. The principal manifestations of the disease involve dilation of the collecting ducts and dysgenesis of the portal triad. Affected fetuses typically display bilaterally enlarged kidneys and oligohydramnios. As a result of the oligohydramnios, fetuses often display the "Potter phenotype" consisting of pulmonary hypoplasia, a characteristic facies, and deformities of the spine and limb (Fig. 3a, b, c and d). Children who survive the neonatal period often die from hypertension, renal failure and portal hypertension due to hepatic fibrosis. It is now known that not all affected individuals die postnatally, but that there is a wide spectrum of phenotypes, including a varied survival rate. However, with increased age there is an increase in life-threatening renal failure (Zerres et al.,1984; Gang and Herrin, 1986; Deget et al., 1995) The frequency of this disease in the general population is estimated to be 1 to 2 for every 10,000 live births (Zeres et al., 1988; Zerres et al., 1996; Zerres et al., 1998; Guay-Woodford et al., 1995). Until quite recently, it was thought that there was only one gene for this disease, PKHD-1, located on chromosome 6p21.1-12 (Fig. 3e). However, recent research has found evidence to refute this belief. Zerres et al. (1999) have found two siblings suspected to be affected with ARPKD but who do not map to the chromosome 6 region. This was the first time that ARPKD-like symptoms did not map to 6p21.1-12 although the phenotype of the affected individuals is described as being similar to six

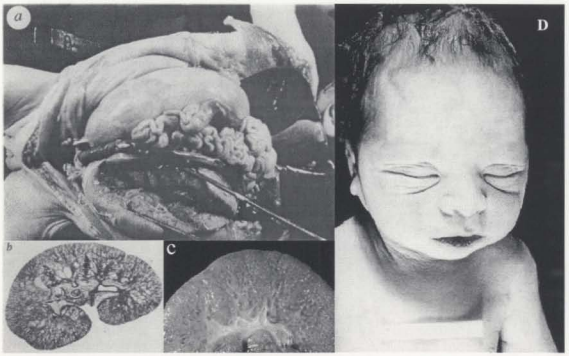
previously described sib pairs (Gillessen-Kaesbach et al., 1993) It is not yet known where the suspected new gene maps.

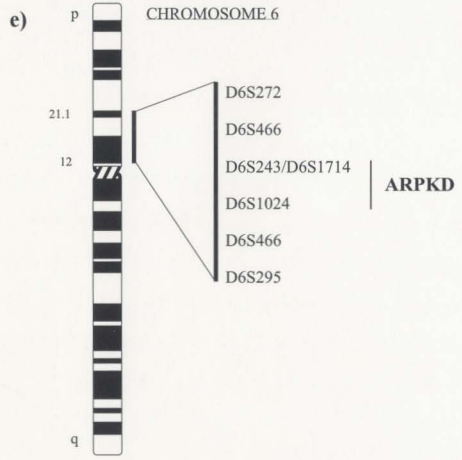
A common feature in the development of ARPKD is the thickening of the collecting tubules caused by localized proliferation and secretion by epithelial cells which develop into fluid filled cysts. These cysts contain biologically active proteins for epidermal growth factor receptor (EGF-R) (Gattone et al., 1990; Avner, 1995). In addition to the mislocalization of EFG-R, it has also been shown that there is an increase in the mRNA, protein and tyrosine kinase activity of EGF-R (Avner and Sweeney, 1995; Wilson et al., 1993; Orellana et al.,1995). Recently, by decreasing the tyrosine kinase activity in mice programmed to develop ARPKD, Richards et al. (1998) were able to show that it is the EFG-R tyrosine kinase activity which causes increased cell proliferation leading to cyst formation. It was also found that Transforming Growth Factor (TGF)-α increased the tyrosine kinase activity of EGF-R and leads to larger cystic lesions. This evidence points to the EGF-TGF-α-EGF-R loop as a mediator of the proliferative expansion of collecting tubule epithelium in ARPKD (Sweeney and Avner, 1998).

1.5 Previous Studies on the Newfoundland Family

Previous study of this family involved microsatellite analysis of the NPH-1 region of chromosome 2 and the ARPKD region of chromosome 6. It was determined that polycystic kidney disease in this family did not map to the NPH-1 locus but that in sibships A and B it did map to the ARPKD critical region of chromosome 6. There was no

Figure 3: Newborn with massively enlarged polycystic kidneys (a) representing severe ARPKD. Cut section reveals multiple cysts giving the kidney a spongy appearance (b) (Zerres et al., 1994). c) Polycystic kidney with uniformly distributed cysts. Note the lack of corticomedullary differentiation. d) Photograph of a newborn displaying the common facial features of Potter phenotype. Note the prominent infraorbital folds. e) Ideogram of chromosome 6 depicting the markers around the ARPKD region.


a
b
c
D


e)
p
CHROMOSOME 6
21.1
12
q
D6S272
D6S466
D6S243/D6S1714
D6S1024
D6S466
D6S295
ARPKD

homozygosity found in the affected individual in sibship C (results reported in detail below) (Jones, 1997).

As there was no of evidence for ARPKD in sibship C, it was decided that additional diseases at other loci should be ruled out before ARPKD could be inferred as the disease affecting the individuals in sibships A and B.

1.6 Objectives of Study

The objectives of this study were to;

i) investigate the possibility of a genetic disorder other than ARPKD in the U1 family.

ii) produce a Phage Artificial Chromosome (PAC) contiguous map of the ARPKD critical region.

iii) find new microsatellite markers in order to narrow the critical region further.

2. MATERIALS AND METHODS

2.1 PREPARING DNA FROM WHOLE BLOOD

Blood was collected in EDTA tubes (volumes are given for 5-7mL of blood). Approximately 45mL of warm (37°C) NH_4Cl:Tris solution (900mL of 0.155M NH_4Cl; 100mL of 0.17 M Tris•HCl, pH 7.65) was added to 5mL of blood in 50mL centrifuge tube and incubated at 37°C for 5 minutes before centrifuging at 2500 rpm (1000 x g) for 5 minutes. The supernatant was poured off, 10mL saline (0.85% NaCl) was added and the tube was vortexed and centrifuged again. The supernatant was poured off and 3mL of nuclei lysis buffer (10 mM Tris•HCl, 400 mM NaCl, 2 mM EDTA, pH 8) was added to the cell pellet which was vortexed and transferred to 15mL centrifuge tube. Then 0.2mL of 10% SDS and 0.5mL of Pronase E solution (3mg/mL in 1% SDS, 2 mM EDTA) was added and the tube was incubated at 37°C overnight (or 2 hours at 55°C).

1mL of saturated NaCl was added and the solution was shaken vigorously for 15 seconds and centrifuged at 2500 rpm (1000 x g) for 15 min. The supernatant was poured into a 15mL tube which was filled with 95% ethanol. The tubes were inverted several times to precipitate out the DNA. The DNA clump was fished out with a with a 9" glass hook. (9" Pasteur pipette; tip melted to form a hook) and washed several times with a stream of 70% ethanol and allowed to air dry before being dissolved in 300-500mL TE (10 mM TRIS•HCl, 1 mM EDTA, pH 8) overnight by cracking off tip of hook in tube and gently mixing on a

rotator for an hour before quantifying using UV spectrometry. DNA samples were stored at 4°C (Based on Miller et al., 1988).

2.2 IDENTIFICATION OF DISEASE LOCI

2.2.1 Screening with Microsatellites

Oligonucleotide primers from the critical regions of NPH-1 (Medhioub et al., 1994), NPH-2 (Haider et al., 1998), mouse bpk/jcpk (Lisa Guay-Woodford personal communication), and ARPKD (Jones, 1997; Mücher et al., 1998) (Appendix 1) were obtained from Research Genetics (Huntsville, AL) and labeled with ^{32}P (Mix: 1.08μL 20μM forward primer, 0.45μL OnePhorAll buffer (Amersham, Bridgewater, NJ), 2.70μL 6000 Ci/mmol γ-^{32}P-ATP (Amersham) and 0.30μL T_4 PNK (Amersham) (total of 4.5μL). Vortex and pulse in a centrifuge then let sit in 37°C water bath for at least 30 min (place in freezer if not using immediately). Add 1μL 1/10 DNA samples to separate small polymerase chain reaction (PCR) tubes. To each tube add 9μL of PCR cocktail (244.88μL dH_2O, 42.2μL 10x Tfl buffer, 5.3μL 25 mM $MgSO_4$, 67.5μL 1.25 mM dNTPs (6.25μL each of 100mM A, C, G and T stock (Amersham) in 475μL dH_2O in a 2mL tube)., 8.78μL each of 20μM forward and 20μM reverse primers, 3.38μL 5u/μL Tfl polymerase (Promega, Madison, WI), 4.5μL radiolabeled primers). The samples were then placed in a Perkin Elmer 9600 thermal cycler and denatured by heating to 95°C for 5 minutes, cycled 35 times using; 95°C for 45 seconds, 45°C - 60°C for 30 seconds, 72°C for 35 seconds, then 72°C for 5 minutes and finally

incubated at 4°C until ready to use. 2.5μL of "DNA Stop" solution" (0.05% bromophenol blue and 0.05% xylene cyanol, 9.5mL formamide and 400μL 0.5M EDTA) was added to the samples and 2.5μL was then run on a 6% polyacrylamide gel (15μL TEMED (Sigma, St Louis, MO), 6mL 10x TBE (108g Trizma base (Sigma), 55g Boric acid (Sigma) and 40mL 0.5M EDTA (BDH Inc, Toronto, ON) in 1L dH_2O), 9mL 40% Acrylamide (dissolve 95g Acrylamide (Bio Rad) and 5g Bis-Acrylamide (Sigma)) in 250mL and filter with Whatman #4, 10mL dH_2O, 20g Urea (Sigma), 19mL Formamide (BDH Inc). 480μL 10% Ammoniumpersulfate (Promega) at 40W constant power. Once the gel had run an appropriate amount of time the electrophoresis was stopped and the gel was transferred to 3mm Whatman paper by drying on a gel drier (Bio Rad, model 583) for approximately 1 ½ hours. Autoradiography was then performed by placing the gel in an X-ray cassette with a DuPont (Mississauga, Ontario) Lightning amplifying screen and X-ray film (AGFA Curix, Toronto, ON) and placed in a dark-room for 1 - 10 days before being developed. Genotyping was performed more than one time when necessary.

2.2.2 Haplotype Analysis

Marker alleles were scored from autoradiographs at least two times by two different individuals in a blinded fashion. Haplotypes were constructed manually for the family for each of the candidate loci using all available microsatellite data in such a way that they represent the minimal number of recombination events. Sharing of haplotypes between affected and unaffected individuals within the family was used to exclude a locus. Support

for linkage to a locus was established if the affected individuals in the U1 family displayed HBD in the critical region and their unaffected siblings did not (Lander and Botstein, 1997) or if affected individuals shared one copy of a suspected mutated locus from one parent and the same chromosomal locus from the other parent. This would reveal the possibility of two separate mutations.

2.2.3 Linkage Analysis

Linkage analysis was carried out in order to support the results of haplotype analysis. Two-point linkage analysis was carried out using the MLINK subroutine of FASTLINK (v4.0P) (Lathrop and Lalouel, 1984; Cottingham et al., 1993; Schaffer et al., 1994). In the case of mouse *bpk/jcpk*, human studies had not previously been performed so full penetrance was assumed. The other diseases were modeled as being recessive with a penetrance of 1 (Haider et al., 1998; Antignac et al., 1993; Zerres et al, 1994). Two point LOD scores were calculated for all microsatellite markers spanning the critical region of each locus. Significantly negative LOD scores ($\leq$-2.0, θ = 0) were used as criteria for exclusion of linkage to a disease gene while positive LOD scores were taken as support for linkage.

2.3 Obtaining PAC Clones

2.3.1 The PAC Libraries

A human PAC screening library was obtained from Genome Systems Inc. (St. Louis,

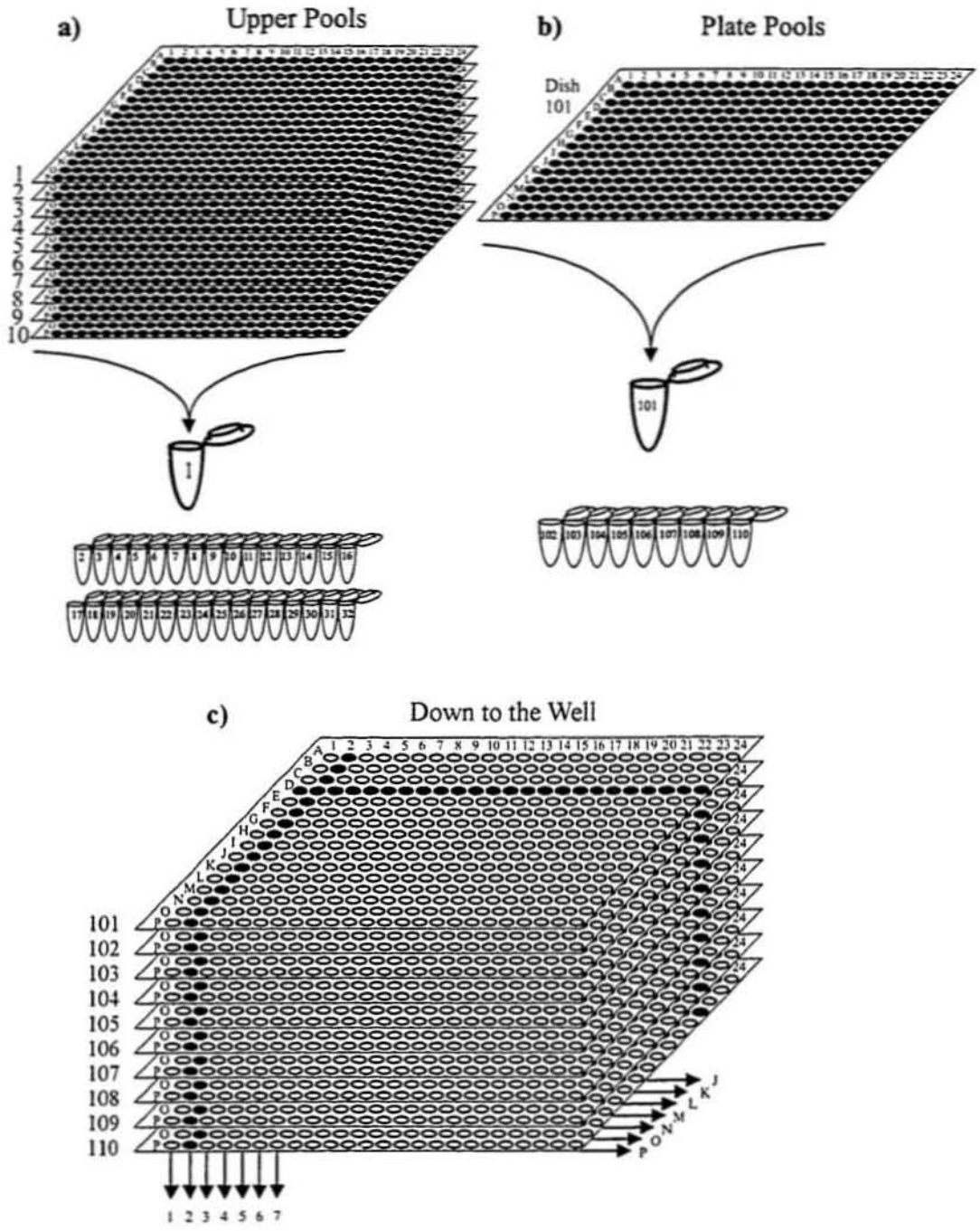
MO) The human PAC library is housed in 321 microtiter dishes, each with 384 wells with one PAC/well. DNA is combined from blocks of 10 dishes for a total of 32 "Upper pools" (i.e., Upper pool #1 has DNA from dishes 1-10). The DNA from each microtiter dish is pooled individually to create 321 "Plate Pools". Once the "Upper Pools" have been identified, a screen of the appropriate plate pools will reveal the plate on which a PAC of interest is located. To create the "Down to the well" pools, DNA from each column and each row is pooled then DNA from 10 column and row pools are combined (Fig. 4).

A PCR is performed for 35 cycles using oligos from the ARPKD region to obtain clones which are able to hybridize to complementary sequences within the region and amplify an STS. The products of the reactions were run on a 2% agrose gel. Clones which contained a sequence complementary to the oligos will display a band identical in size to a positive control. By using this method it was possible to narrow PAC clones to a specific dish, column, and row. Once PACs were identified they were subject to PCR using other markers within the region in order to determine which marker sequences they contained.

2.3.2 Isolating PAC DNA

PAC clone agar stabs were ordered from Genome Systems Inc.. Bacteria from the stabs were streaked on 25μg/mL LB-kanamycin (LB-kan) plates and incubated at 37°C overnight. The next day a single colony was picked and transferred to a 25mL flask containing 3mL 25μg/mL LB-kan and incubated at 37°C overnight with 250rpm shaking. The next morning 2mL of the overnight culture was used to innoculate 500mL of 25μg/mL

Figure 4: a) Illustration depicting how the PAC "Upper Pools" are created. The picture indicates that there are 32 upper pool samples, each containing the DNA of 10 plates which is equal to 3840 PAC clones. b) Illustration depicting how the PAC "Plate Pools" are created. The DNA from the 384 PACs on each dish are combined to create one sample. c) Illustration depicting how the PAC "Down to the Well" pools are created. Each column and each row of a group of 10 plates are combined into one sample.


a)
Upper Pools
b)
Plate Pools
Dish
101
c)
Down to the Well
101
102
103
104
105
106
107
108
109
110
1 2 3 4 5 6 7

LB-kan in a 2L flask. The culture was incubated at 37°C with 250rpm shaking for 1 ½ hours. The culture was then induced with IPTG to a final concentration of 0.5mM and incubated, as above, for a further 5 hours. The culture was then centrifuged in a 250mL polypropylene bottle at 5000xg for 10 min. and the supernatant was poured off. At this point the bacteria may or may not have been stored overnight at -20°C and thawed on ice when needed. The cell pellet was resuspended in 20mL P1 buffer (50mM Tris•Cl, pH 8.0, 10mM EDTA, 100μg/mL RNase AI) and lysed by adding 20mL P2 buffer (200mM NaOH, 1% SDS (Sigma, St. Louis, MO) made fresh daily) and incubated at room temperature for no more than 5 min. The lipids, proteins and SDS were then precipitated by adding 20mL neutralization buffer P3 (3.0M KAc, pH5.5, stored at 4°C) and incubating on ice for 15 min. The cellular material was then transferred to two 48mL polypropylene tubes and centrifuged at 10 000 rpm for 30 min. The supernatant was recentrifuged, if needed, and transferred to a fresh 50mL tube. A Qiagen (Valencia, California) tip 500 was equilibrated with 10mL QBT buffer (750nM NaCl, 50mM MOPS (Sigma), pH 7.0, 15% isopropanol, 0.15% Triton-X (Beckman Instruments, Fullerton, California) and allowed to run through before the cell supernatant was added to the tip 500. Once all the cell lysate had run through, the tip was washed with 30mL buffer A (0.75M NaCl, 50mM MOPS, pH 7.5, 15% ethanol) and eluted into a 48mL polypropylene tube using two 5mL aliquots of buffer B (1.4M NaCl, 50mM Tris•HCl, 15% ethanol, pH 8.1-8.2) heated to 65°C (Appendix 2). The DNA was then precipitated and resuspended in 500μL TE buffer (10mM Tris•HCl, 1mM EDTA, pH8.0). The DNA was quantified using a fluorimeter and stored at 4°C. The method usually

produced a yield of 30-60μg of PAC DNA.

2.4 PRECIPITATING AND QUANTIFYING DNA

In order to sequence the ends of the PAC clones it was necessary to first determine the mass of DNA obtained and precipitate a specific aliquot. DNA was quantified using a DNA fluorometer TKO 100 (Hoefer Scientific Instruments, San Francisco). 2mL fluorometer buffer (10mL TNE (100mL 1M Tris•HCl (pH 7.4), 100mL 0.1M EDTA, 58.4g NaCl in 1000mL dH_2O), 90mL dH_2O, 10μL 1mg/mL Hoechst dye) was placed into a cuvette, the "scale" knob was turned all the way to the right and the machine was zeroed using the "zero" knob. 4mL of 250μg/mL calf thymus DNA was added to the cuvette and mixed gently and the scale knob was adjusted to obtain a reading of 500. The cuvette was poured out and rinsed with 1mL buffer before adding 2mL buffer which was used to rezero the machine. 5mL of sample DNA was added to the cuvette, mixed, placed in the fluorometer and the scale reading was taken. The concentration of DNA in the undiluted sample was taken to be 2 X scale reading/5 (μg/mL).

DNA was precipitated by adding 10% 2M NaAc (pH 3) and two volumes of -20°C 95% ethanol and mixing. The mixture was then placed at -20°C overnight or at -70°C for 15-20 minutes and spun in a microfuge for 20 minutes. The supernatant was poured off and the sample was washed with 70% ethanol and spun for 10 minutes. The ethanol was wicked off and the DNA pellet was dried under vacuum at room temperature.

2.5 Producing Non-commercially available Oligonucleotide Primers

Primers for which the sequence was known but which were not commercially available due to low heterozygosity or because they were newly discovered were inputted into the Cortec primer design automated order form (http://www.cortec.queensu.ca/). The web based "primer3" (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) was used to design novel primers from newly discovered DNA sequences. This program automatically designed primers around a specified sequence, automatically picking oligos with optimal traits. These primer sequences were then inputted into the Cortec primer design page. All primers were synthesized on a scale of 0.2μmole and were TLC purified.

2.6 Subcloning PAC Clones

In order to screen the PAC clones for novel microsatellite repeats it was necessary to cut the DNA into smaller pieces, ligate it into a suitable vector and transform it into a suitable host bacterial cell to be screened for repeat DNA.

2.6.1 DNA Preparation

16μg of pUC18 DNA (a kind gift from Donna Jackman who was working in Dr. David Heeley's lab) was digested with 2μL (20,000u/mL) BamHI (New England Biolabs,

Beverly, MA) using 2x OnePhorAll buffer (Amersham) and dH_2O in 40μL total volume and incubated in a 37°C water bath overnight. The enzyme was deactivated by placing the reaction in a 65°C water bath for 15 minutes. Digested 10μg of PAC DNA with 1.5μL (6000u/mL) Sau3A (Amersham) using 1x OnePhorAll buffer (Amersham) and dH_2O in 50μL total volume and incubated in a 37°C water bath overnight. The enzyme was deactivated by placing the reaction in a 65°C water bath for 15 minutes. To assess the digestion, 4μg of the digested pUC18 and 5μg of the digested PAC DNA were run a on a 1% agrose gel beside a λ-HindIII marker (Amersham).

1μL (1u/μL) Shrimp Alkaline Phosphatase (SAP) (United States Biochemical Corporation, Cleveland, OH) was added to the remaining 30μL pUC18 cocktail along with 3.5μL 10x SAP buffer. The mixture was incubated in a 37°C water bath for 1 hour. The enzyme was deactivated by placing the reaction in a 65°C water bath for 15 min.

In order to minimize the amount of bacterial chromosomal DNA present in the PAC DNA, each of the isolated PAC samples were treated with 15u of 10u/μL Plasmid Safe adenosine triphosphate (ATP)-Dependent DNase (Epicentre Technologies, Madison, WI) and ATP (Epicentre Technologies) to a final concentration of 1mM. The reactions were placed in a 37°C water bath and after one hour, ATP was again added to a final concentration of 2mM and the reaction was allowed to continue overnight. The DNA was reisolated from the reaction by performing a 1:1 Phenol (Fisher Scientific, Fair Lawn, NJ)-Chloroform (Fisher Scientific) extraction followed by a Chloroform extraction alone. The DNA was then precipitated and washed several times with 70% ethanol. The DNA samples were

resuspended to a final concentration of 0.1μg/μL.

2.6.2 DNA Ligation

1μL (0.353μg) pUC18 (digested and treated with phosphatase) was combined with 0.56μL (0.106μg) digested PAC DNA, 1μL ligation buffer (660μL1M Tris, 66μL 1M $MgCl_2$, 100μL 1M DDT, 50μL 100mM ATP and 124μL dH_2O split into 50μL aliquots and stored at -20°C), 1μL T_4 DNA ligase (MBI Fermentas, Flamborough, ON) and 6.44μL dH_2O, for a total reaction volume of 10μL. A pUC18 negative control (1μL pUC18, 1μL Ligation buffer, 1μL T_4 DNA ligase, 7μL dH_2O) was also set up for comparison. The ligation reactions were allowed to incubate at room temperature overnight.

2.6.3 Preparing Competent cells

E. coli DH2 cells were grown at 37°C overnight in 3mL of LB broth and 0.5mL of that culture was used to innoculate 500mL LB broth the next day. The cells were grown at 37°C until the culture reached an optical density of 0.800 - 1.000. The cells were centrifuged in a polypropylene bottle at 4000 rpm for 10 minutes and the supernatant was poured off. The cells were resuspended in 250mL 0.1M $CaCl_2$ and incubated on ice for 30 minutes. The cells were recentrifuged and resuspended in 5mL cold 0.1M $CaCl_2$ + 15% glycerol and split into 50μL aliquots and frozen at -70°C until ready for use.

2.6.4 Cell Transformation

A 50μL aliquot of competent *E. coli* DH2 cells was thawed on ice and 450μL cold 0.1M $CaCl_2$ was added. In a 1.5mL tube, 80μL *E. coli* DH2 cell suspension was added to the10μL ligation reaction and incubated at room temperature for 30 min. The cells were heat shocked by placing them in a 42°C water bath for no more than 45 seconds. 1mL LB was added and placed in a 37°C water bath for 40 minutes. Cells were then centrifuged cells in a microfuge for 1 minute, the supernatant was poured off and the tube was gently tapped to resuspend the cells in the leftover supernatant. Each suspension was divided in half and spread on 50 μg/mL LB-Carbenicillin (LB-carb) plates and placed in a 37°C incubator overnight.

2.7 Screening for Novel Microsatellites

2.7.1 Transferring Colonies and Fixing DNA

Colonies were picked and transferred, in duplicate, to gridded LB-varb plates which were incubated at 37°C overnight. The colonies were transferred to nitrocellulose paper (Gelman Science, Ann Arbor, MI), placed in a glass dish, colony side up, containing filter paper soaked with 10% SDS and incubated for 3 minutes. The nitrocellulose was then placed in a second dish containing filter paper soaked in denaturing buffer (0.5M NaOH, 1.5M NaCl) and incubated for 5 minutes. The nitrocellulose was then placed in a third dish containing filter paper soaked in neutralization buffer (1.5M NaCl, 0.5M Tris•HCl pH 8.0)

and incubated for 5 minutes. Finally the nitrocellulose was transferred to a fourth dish containing filter paper soaked with 2x SSPE buffer (20x SSPE; 3.6M NaCl, 200mM NaH_2PO_4 and 20mM EDTA pH 7.4) where the cell debris was rubbed off with a tissue. The nitrocellulose was incubated in that solution for 5 minutes before being rinsed in 2x SSPE and place on filter paper and allowed to dry overnight at room temperature.

2.7.2 Probing Colonies for STRPs

Nitrocellulose membranes were placed in glass hybridizaton tubes with 15mL of prehybridization solution (25mL 20x SSPE, 10mL 50x Denhardt's solution (5g ficol, 5g polyvinylpyrrolidone, 5g bovine serum albumin) 0.1%SDS in 500mL, filtered and stored at -20°C), and rotated in a hybridization oven (Robbins Scientific, model 400), set at 50°C, for 2 hours. Three oligos, $(GT)_{10}$, $(CT)_{10}$ and $(CAG)_5$, were radiolabeled with ^{32}P by combining 1.5μL of each 10μM oligo, 4.5μL 6000 Ci/mmol γ-^{32}P-ATP, 1μL 10x T_4 polynucleotide kinase (PNK) buffer, 1μL T_4 PNK and incubating at 37°C for 30 minutes. Once the 2 hour incubation was complete the prehybridization solution was removed and replaced with fresh prehybridization solution containing 5μL of each of the three oligos and rotated in the hybridization oven overnight.

The next day the membranes were removed from the hybridization tubes and washed three times in 200mL of fresh 2x SSPE in a glass dish with gentle shaking for one hour each wash. The membranes were then allowed to air dry and placed in an X-ray cassette with X-ray film for three nights with a Lightning-Plus amplifying screen (DuPont Cronex), at -70°C

before developing.

2.8 SOUTHERN BLOTTING

In order to confirm that the suspected subclones did actually contain microsatellite DNA, the individual subclones were grown up in order to obtain a large quantity of subclone DNA. The PAC fragment insert was then digested, run on a gel, southern blotted, and probed to determine if the oligo properly hybridized to a repeat sequence within the PAC subclone fragment.

2.8.1 Isolating subcloned DNA

Bacteria containing subclones suspected of having microsatellite repeats were grown overnight in 3mL of LB broth in a 25mL flask at 37°C. The bacteria were then isolated using the Wizard miniprep (Promega) system according to the manufacturer's protocol and resuspended in 50μL TE buffer.

2.8.2 Analyzing Subclones

Once isolated, the DNA insert was cut out of the pUC18 vector by digesting the DNA sample in the following reaction; 1μL 20,000u/mL EcoRI (Amersham), 1μL 18,000u/mL HindIII (Amersham), 8μL DNA, 4μL OnePhorAll buffer (Amersham), 6μL dH_2O. The reaction was placed in a 37°C water bath overnight before being run on a 1% agrose (Sigma)

gel for 3 hours.

2.8.3 Transferring DNA

Once properly photographed, the gel was placed in a glass dish containing 500mL denaturing solution (0.6M NaCl, 0.4M NaOH) with gentle agitation for 20 minutes at room temperature. The solution was changed and the incubation was repeated once more. The gel was then incubated in 500mL neutralizing buffer (1.5M NaCl, 0.5M Tris•HCl, pH 7.5) for 20 minutes. Again, this was done with gentle agitation at room temperature and repeated one time. The gel was then placed in 10x SSC (20x SSC: 3M NaCl, 0.3M Sodium Citrate [Trisodium Salt Dihydrate]) at room temperature with gentle agitation for 15 minutes. A Gene Screen Plus hybridization membrane(NEN Life Science Products, Boston, MA), was cut to the size of the gel and placed in water to wet and then into a 10x solution of SSC for 15 minutes. A glass plate was placed over a glass dish containing 200mL 10x SSC and a piece of 3mm chromatography paper (Whatman, Clifton, NJ) was used as a wick by cutting it to the width of the gel and placing it over the plate with each end in the 10x SSC. The gel was placed on top of the wick and Parafilm (American National Can., Menasha, WI) was placed over any part of the wick that the gel did not cover. On top of the gel was placed the transfer membrane and a piece of 3mm Whatman paper (cut to size). On top of this was placed a stack of paper towels approximately six to eight inches thick, and a glass dish for weight. The DNA was then allowed to transfer for 16 to 24 hours.

2.8.4 Probing for STRPs

The Southern blot membrane was probed in the same manner as the bacterial subclone colonies in section 2.7.2. Once probed the membrane was washed three times in 250mL 2xSSC, the first wash was done in the hybridization tube in the hybridization oven at 50°C while the next two washes were done in a glass dish with gentle agitation at room temperature. The membranes were then allowed to air dry and placed in an X-ray cassette with X-ray film for three nights at room temperature before developing.

Any positive clones were picked and grown in 10mL LB-carb as described above. The plasmid DNA was isolated using the Wizard miniprep system according to the manufacturer's protocol and the samples were resuspended in 50μL of water and sent to the University of Guelph for automated sequencing.

2.9 DNA SEQUENCING

Sequencing was performed in order to identify and examine segments of interest such as PAC ends and subclone microsatellites. Sequences were often used to design oligonucleotide primers.

2.9.1 Cycle sequencing

All reagents used for cycle sequencing were from the "fmol sequencing kit" purchased from Promega (Madison, WI). An oligonucleotide primer was radioactively

labeled by combining 10pM of primer, 1μL 10x T4 PNK buffer, 5 units (u) T4 PNK and 3μL 10mCi/mL γ-^{32}P-ATP in 10μL total volume and incubating at 37°C for 30 min. For each sample, 4 tubes were labelled A, C, G and T and to each 2μL of the appropriate d/ddNTP was added. For each sample a sequencing cocktail of 5μL 5x sequencing buffer, an appropriate mass of DNA template, 1.5μL end labeled primer and 1μL of 5 u/μL sequencing grade Taq polymerase in a final volume of 17μL was created. 4μL of the sequencing cocktail was added to each of the tubes marked A, C, G and T and a drop of mineral oil was added before capping. The samples were placed in a Perkin Elmer 9600 thermal cycler and denatured by heating 95°C for 2 min. and cycled at 95°C - 45sec., 50°C - 30sec., 70°C - 1min., 35 times and then ramped to 4°C until ready for use. 3μL of "DNA Stop" solution was added to each sample and 2.5-3μL was then loaded onto a urea/polyacrylamide gel and run for 1½ and 3 hours.

2.9.2 Automated sequencing

A 1.25μg sample of PAC DNA was precipitated and used in the following reaction; 8.5μL dH2O, 1μL DMSO, 25pM primer and 8μL ABI BigDye solution (Perken-Elmer Applied Biosystems, Foster City, CA). The samples were placed in a Perkin Elmer 9600 thermal cycler and denatured by heating 95°C for 5 min. and cycled at 95°C - 45sec., 50°C - 20sec., 60°C - 4min., 55 times and then ramped to 4°C until ready for use. The samples were then purified (Helix and Primer, St. John's, NF) and run for 12 hours in an ABI Applied Biosystems 370A automated sequencer using a 30w Helium-Argon laser. The gel was run

for 12 hours and the output was analyzed using ABI Sequence analysis software version 1.2. Alternatively, the DNA samples were dried down in a 1.5mL tube and sent to the Department of Zoology at the University of Guelph for automated sequencing. Sequencing in this manner routinely resulted in over 600 bases of readable sequence.

2.10 Analyzing DNA Sequences

Clean DNA sequences were typed into The National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) online advanced BLAST search engine. The searches used the BLASTN program with the nr database. Any sequences which were found to be related to the query sequence with an E-value of 0.01 or lower were considered significant.

In order to search for repeat sequences, the program "Sputnik" was used (http://www.abajian.com/sputnik/index.html). Sputnik searches DNA sequence files in FASTA format for microsatellite repeats with a repeated region between 2 and 5 base pairs. The program reads through the entire sequence and assumes the existence of a repeat at every position and compares subsequent nucleotides. If the resulting score rises above a preset threshold, the region along with its position and score is written out. If the score falls below a cutoff threshold, the search is abandoned and begun again at the next nucleotide. Each nucleotide that matches the value predicted (by assuming a repeat) adds to the score whereas each "error" subtracts from the score. The FASTA format sequence was obtained

from NCBI and saved as a text file in UNIX format before being analyzed by Sputnik.

3. IDENTIFICATION OF DISEASE

3.1 JUVENILE NEPHRONOPHTHISIS I (NPH-1)

3.1.1 Results and Discussion

As seen in Fig. 5, the affected individuals in the U1 family did not display HBD for any of the polymorphic markers within the NPH-1 region nor was there any evidence of a shared haplotype. The LOD scores significantly ruled out four of the markers and indicated that it was improbable for D2S121to be linked to the NPH-1 locus due to the low negative score of -1.41 at $\theta = 0$ (Table 2). This finding indicated that none of the affected individuals inherited two mutated copies of the NPH-1 and the disease of the U1 family is not linked to NPH-1 (Jones, 1997).

Figure 5: Core pedigree of the U1 family with all affected individuals, their parents, grandparents and other select relatives depicting their haplotypes for the markers within the NPH-1 region of chromosome 2. Note that the affected individuals are not homozygus for any of the markers and have all inherited different chromosomes. [] = inferred genotype, - = unknown genotype, x = meiotic recombination.

Fig 5. U1 NPH-1 haplotypes

Table 2: Two point LOD scores between the NPH-1 locus and markers in the NPH-1 region of chromosome 2*.

Marker	$\theta = 0.00$	$\theta = 0.05$	$\theta = 0.10$
D2S293	$-\infty$	-2.37	-1.31
D2S340	$-\infty$	-2.33	-1.20
D2S160	$-\infty$	-3.02	-2.24
D2S121	-1.41	0.76	0.93
D2S363	-5.68	-1.58	-0.074

* Reproduced from Jones (1997).
θ Recombination frequency

3.2 JUVENILE NEPHRONOPHTHISIS II (NPH-2)

3.2.1 Results and Discussion

As seen in Fig. 6, the haplotypes for the family, using the chromosome 9 NPH-2 markers, did not provide any evidence to indicate HBD among the affected individuals at any loci. The affected individuals in sibship A were not exclusively homozygous at any of the 6 loci tested except for the possible exception of D9S280 which is, in fact, a flanking marker. A recombination in individual 51 created a heterozygosity at marker D9S910 whereas her siblings are homozygous, thus ruling out that locus. Similarly, there are no common homozygous haplotypes among the affected individuals of sibship B or C. LOD score analysis confirmed that the disease was not linked to NPH-2 as all of the scores were below -2 at $\theta = 0$ (Table 3).

Figure 6: Core pedigree of the U1 family with all affected individuals, their parents, grandparents and other select relatives depicting their haplotypes for the markers within the NPH-2 region of chromosome 9. Note that the affected individuals are not HBD for any of the marker loci. [ ] = inferred genotype, - = unknown genotype, x = meiotic recombination.

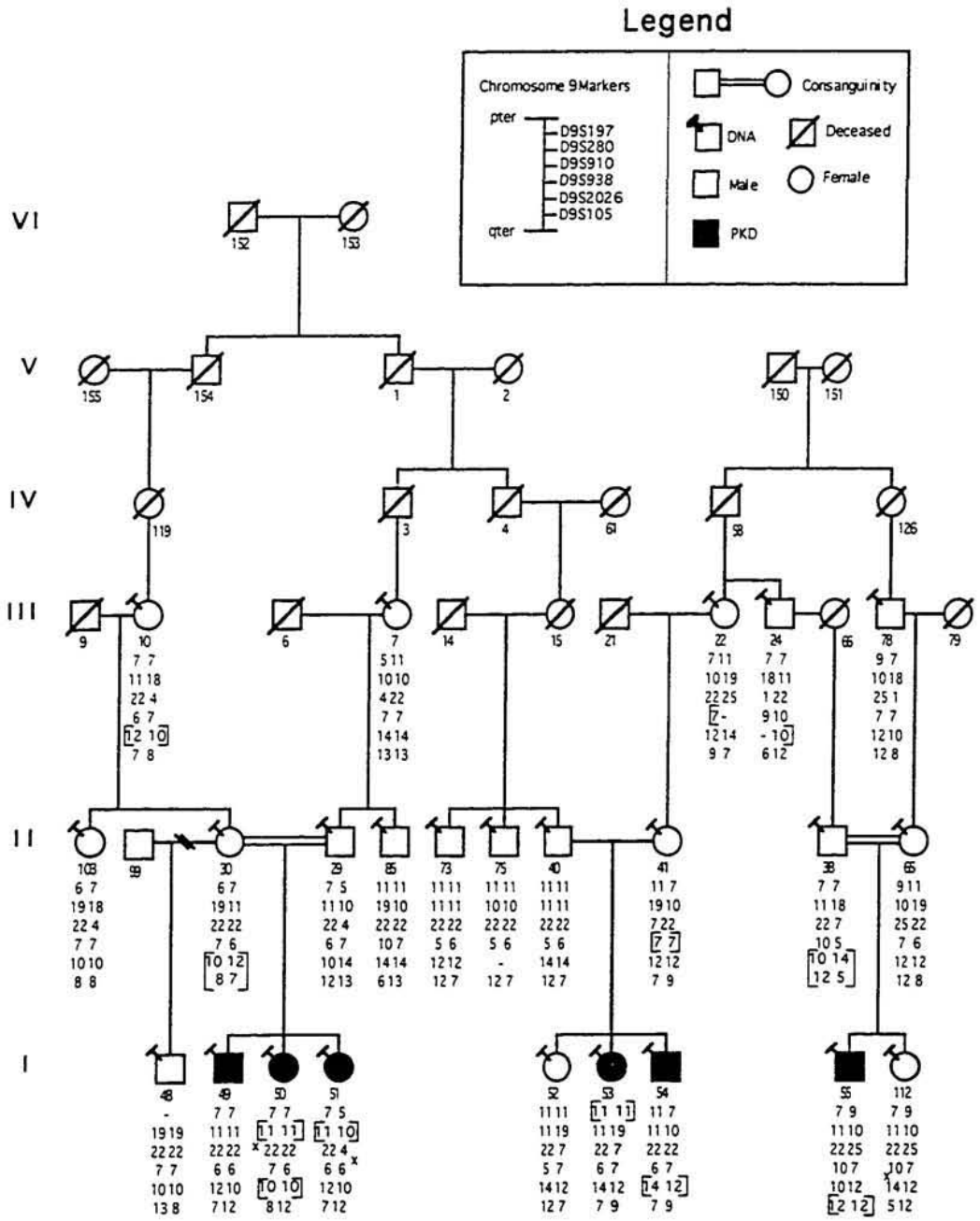

Legend
Chromosome 9 Markers
pter
D9S197
D9S280
D9S910
D9S938
D9S2026
D9S105
qter
Consanguinity
DNA
Deceased
Male
Female
PKD
VI
V
IV
III
II
I

Table 3: Two-point LOD scored between the NPH-2 locus and markers in Fig. 6.

Marker	$\theta = 0.00$	$\theta = 0.05$	$\theta = 0.10$
D9S197	$-\infty$	-1.88	-0.900
D9S280	$-\infty$	-2.53	-1.53
D9S910	$-\infty$	-1.60	-0.82
D9S938	$-\infty$	-1.07	-0.34
D9S2026	-3.37	-1.82	-1.21
D9S105	$-\infty$	-0.96	-0.76

θ Recombination frequency

3.3 MOUSE BPK/JCPK

3.3.1 Results and Discussion

Polymorphic markers for this part of the study were a kind gift from Lisa Guay-Woodford, Department of Medicine, University of Alabama at Birmingham, Birmingham. As seen in Fig. 7, none of the affected individuals displayed HBD at any of the five markers tested. Two point linkage analysis statistically excluded linkage to the *bpk/jcpk* locus for all five of the markers tested as all gave scores of $-\infty$ at $\theta = 0$ (Table 4). These results effectively revealed that the kidney pathology occurring in the U1 family was not linked to a human homolog of the mouse *bpk/jcpk* gene.

Figure 7: Core pedigree of the U1 family with all affected individuals, their parents, grandparents and other select relatives depicting their haplotypes for the markers centenic to the *bpk/jcpk* region of chromosome 10. Note that the affected individuals are not HBD for any of the marker loci along the chromosome. [ ] = inferred genotype, - = unknown genotype, x = meiotic recombination.

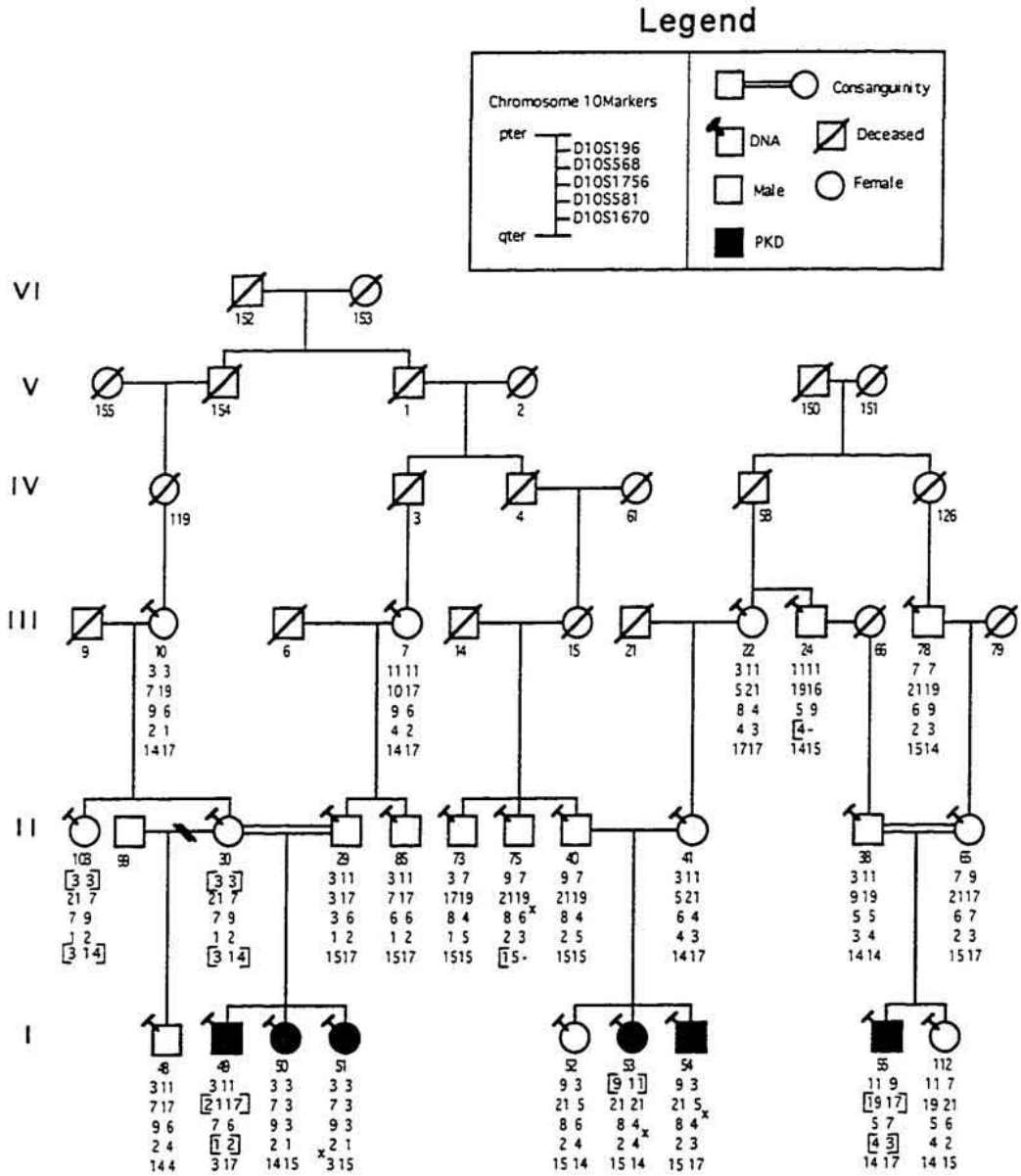

Legend
Chromosome 10Markers
pter
D10S196
D10S568
D10S1756
D10S581
D10S1670
qter
Consanguinity
DNA
Deceased
Male
Female
PKD
VI
V
IV
III
II
I

Table 4: LOD scores of two-point linkage analysis on the markers in the bpk/jcpk region.

Marker	$\theta = 0.00$	$\theta = 0.05$	$\theta = 0.10$
D10S196	$-\infty$	-4.18	-2.82
D10S568	$-\infty$	-3.25	-2.27
D10S1756	$-\infty$	-4.12	-2.62
D10S581	$-\infty$	0.86	1.02
D10S1670	$-\infty$	-6.15	-4.12

θ Recombination frequency

3.4 Autosomal Recessive Polycystic Kidney Disease

3.4.1 Results and Discussion

As seen in Fig. 8, all individuals in sibship A are HBD for markers D6S272, D6S465, D6S243 and D6S1714, revealing a common homozygous 5-5-12-5 haplotype. However, due to a meiotic recombination in individual PID 29, the affected individuals are heterozygous at the marker loci D6S1024, D6S466 and D6S295. Both affected individuals in sibship B (PID 40 and 41) are homozygous for the D6S272 marker and are heterozygous for the 5-5-12-5 haplotype inherited from their father (PID 40) which is also found in sibship A. Both individuals also share the same maternal chromosome throughout the full region. Their unaffected sister (PID 52) does not share either of the chromosomal haplotypes with her affected siblings. The affected individual in sibship C (PID 55) is heterozygous for all markers tested and does not have the 5-5-12-5 haplotype seen in the other two sibships. His unaffected sister (PID 112) was unable to be tested for much of the study due to her young age; however, the marker D6S1714 was used to examine her DNA and based on the results it appears that she and her brother inherited different chromosomes from each parent. For sibships A and B, two-point linkage analysis revealed several positive LOD scores, including two which indicated significant linkage between the disease in the family and markers at the PKHD-1 locus (LOD >2.0 at $\theta = 0$) (Table 5). These results support the possibility that the disease affecting the U1 family is ARPKD. However, the results do not conclusively reveal that ARPKD is the disease in all affected individuals.

Figure 8: Core pedigree of the U1 family with all affected individuals, their parents, grandparents and other select relatives depicting their haplotypes for the markers in the ARPKD region of chromosome 6. Note that the affected individuals in sibship A are HBD for the markers D6S272 to D6S1714 creating a 5-5-12-5 haplotype. Affected individuals in sibship B have one copy of the 5-5-12-5 haplotype and have the same copy of the maternal chromosome which they do not share with their unaffected sister. The affected individual in sibship C is not HBD for any of the marker loci within the ARPKD region. [ ] = inferred genotype, - = unknown genotype, x = meiotic recombination.

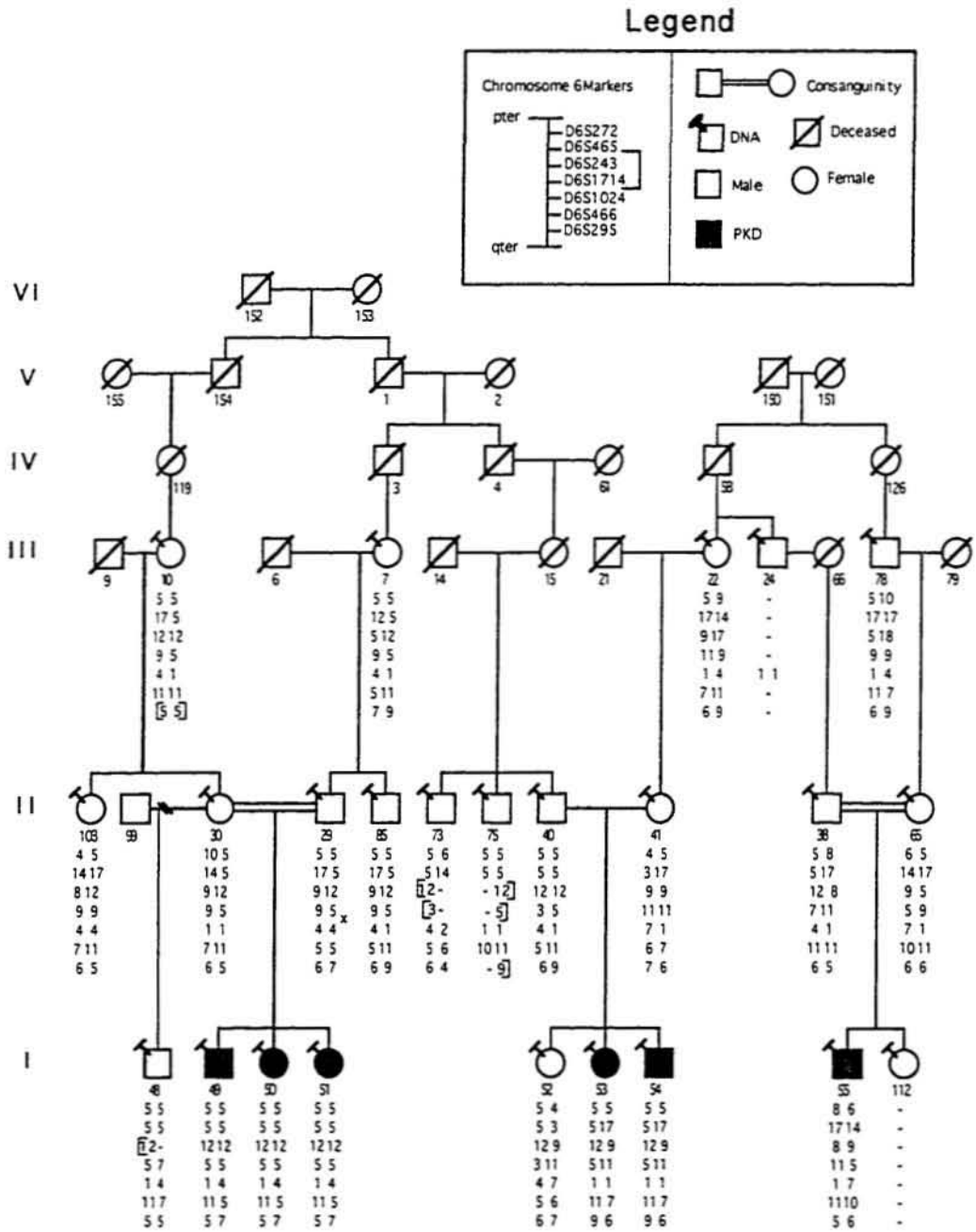

Legend
Chromosome 6 Markers
pter
D6S272
D6S465
D6S243
D6S1714
D6S1024
D6S466
D6S295
qter
Consanguinity
DNA
Deceased
Male
Female
PKD
VI
V
IV
III
II
I

Table 5: LOD scores of two-point linkage analysis on the markers in the ARPKD region.

Marker	$\theta = 0$	$\theta = 0.05$	$\theta = 0.10$
D6S272	-1.48	1.69	1.83
D6S465	2.95	3.98	3.70
D6S243	-0.18	1.26	1.50
D6S1714	3.08	2.67	2.38
D6S1024	-1.41	1.10	1.23
D6S466	-3.85	0.90	1.33
D6S295	-2.74	1.44	1.77

* From Jones (1997)
θ Recombination frequency

It does appear that the affected individuals in sibship A (PIDs 49, 50 and 51) have inherited two identical copies of the mutated PKHD-1 gene and that the affected individuals in sibship B (PIDs 53 and 54) have likely inherited two different mutated copies of PKHD-1. However, the affected individual in sibship C (PID 55) did not inherit the 5-5-12-5 haplotype present in the other affected individuals or the 5-17-9-11-1-7-6 haplotype present in the affected individuals in sibship B. This is consistent with the clinical phenotype which suggests that the affected individual in sibship C has a different disease from sibships A and B. It remains unclear as to what disease individual PID 55 has. Assuming that the disease in sibships A and B is indeed ARPKD, the meiotic recombination in individual PID 29 narrows the ARPKD critical region from flanking markers D6S1714/D6S243 (Mücher et al., 1994)-D6S466 (Guay-Woodford et al., 1995) to flanking markers D6S1714/D6S243-D6S1024 creating a 1cM critical region. These results were later corroborated by Zerres et al. (1998) who identified a meiotic recombination in an individual, allowing the confirmation of D6S1024 as the centromeric flanking marker for the ARPKD critical region.

3.5 Autosomal Dominant Polycystic Kidney Disease

3.5.1 Results and Discussion

As has been noted numerous times in the available literature, the most effective method of determining if an individual has ADPKD is to perform an ultrasound on the parents of an affected child (Ogborn, 1994; Fick and Gabow, 1994; Sessa et al., 1997; Zerres

et al., 1998). By doing so it can be determined if one or both parents have polycystic kidneys. If at least one parent has polycystic kidneys a diagnosis of ADPKD can be made. Kidney ultrasounds were carried out by nephrologist Dr. Benvon Cramer (MD), Department of Medicine, Health Sciences Centre, St. John's, Newfoundland. The only individual for which there was any suggestion of ADPKD was PID 55 in sibship C as he did not apparently map to any of the other regions. The father, PID 38, and mother, PID 65, underwent kidney ultrasound in order to determine if either of them had polycystic kidneys. The father had a small cyst at the corticomedullary junction of his left kidney. Otherwise, his liver and both his kidneys were normal. The mother had normal kidneys and liver. The finding that the father of the affected individual had a cyst on his kidney was equivocal. In ADPKD 1, cysts develop in both kidneys by age 30 although the onset of cystogenesis may occur later in ADPKD 2 (Patrick S. Parfrey, personal communication). Therefore, it is likely that ADPKD was not inherited by PID 55 from either of his parents; however, this does not rule out the possibility of the presence of a novel mutation in this individual.

3.6 Conclusion of Disease Identification

Homozygosity mapping along with linkage analysis on the U1 family using markers in the NPH-1, NPH-2 and *bpk/jcpk* regions allowed these diseases to be ruled out as the cause of PKD in this family. These techniques also revealed that the ARPKD is the probable cause of the PKD in sibships A and B but not sibship C. Further analysis of sibship C failed

to reveal evidence of ADPKD. This unfortunately means that the cause of the kidney pathology in sibship C remains unknown. It is possible that an ADPKD mutation in sibship C occurred spontaneously in individual PID 55. Although rare, there are other examples of several mutations leading to similar diseases occurring in the same family as was the case for a Newfoundland family displaying hereditary eye disease. Because there are other individuals with ARPKD within the Newfoundland population, it is likely that two separate ARPKD mutations are causing the pathology in sibship B (Jane Green, personal communication). Sibships A and B in the U1 family are informative and allowed the ARPKD critical region to be narrowed to approximately a 1cM region. Because of this, the family was considered potentially useful for further narrowing the ARPKD critical region.

4. PRODUCTION OF A PAC CONTIGUOUS MAP FOR THE ARPKD CRITICAL REGION

4.1 SCREENING THE PAC "DOWN TO THE WELL" LIBRARY POOLS

In 1997, Lens et al. published the results of a fine physical map of the ARPKD region, made up of novel sequence tagged site (STS) markers and a YAC contig (Fig. 9). The markers specified in that study were synthesized and used to screen the PAC screening library in order to obtain a more detailed map of the region.

4.1.1 Initial Screen

A total of seven PACs were identified which overlapped the ARPKD region (Table 6, Fig. 10, 11 and 12); however, they did not completely cover the whole region. The ends of some of the PACs were sequenced using the T7 or SP6 primer (Appendix 3 and 4) and subject to a BLASTN search (Appendix 5). The sequences attained were used to produce STS markers which enabled proper orientation of the clones and further screening (Appendix 6).

4.1.2 Second Screen

The new markers obtained from the first seven PAC clones were used to rescreen the

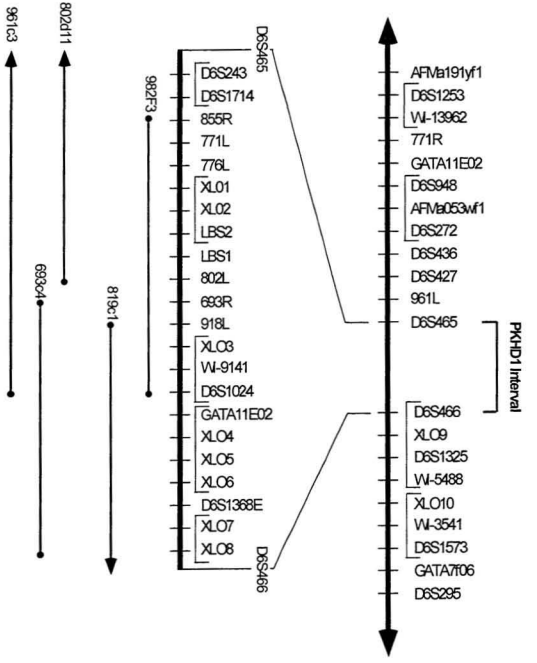


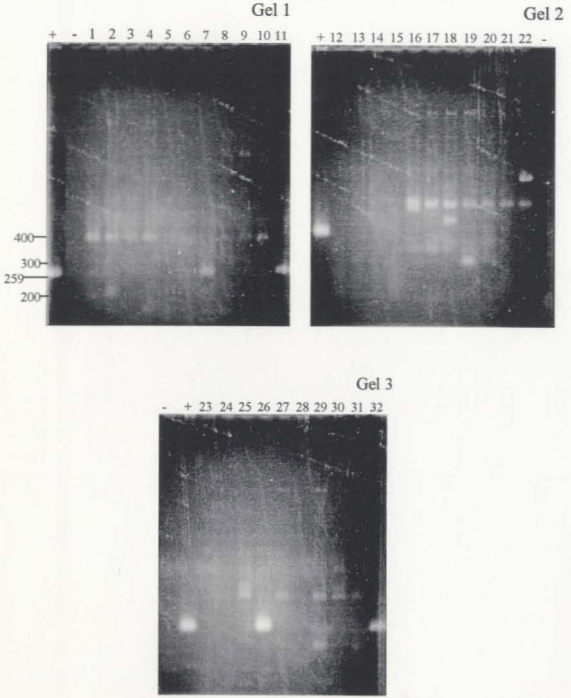
Figure 9: Fine Map of the ARPKD region with new nonpolymorphic STSs and five YAC clones covering the region (Lens et al., 1997).

Table 6: Results of initial screen of PAC library using polymorphic markers and STSs.

Marker*	Annealing temperature (°C)*	# of possible PACs Initially Found	Final # of PACs Found Accurately	PACs Identified
D6S1714	60	4	1	123h3
D6S1024	45	3	1	122k3
776L	58	4	2	101d2, 320c12
LBS1	55	2	2	203o6, 265a14
XL03	58	2	1	146o10

* From Lens et al. (1997)

Figure 10: PAC "Upper Pool" screening using the marker 855R. The expected product was 259bp in size which appeared in lanes 7, 11, 26, and 32. The table below the gels indicates which plates have been pooled to create each sample. + = Positive control (human genomic DNA), - = Negative control (dH_2O).



1. 1-10	7. 61-70	13. 121-130	19. 181-190	25. 241-250	31. 301-310
2. 11-20	8. 71-80	14. 131-140	20. 191-200	26. 251-260	32. 311-321
3. 21-30	9. 81-90	15. 141-150	21. 201-210	27. 261-270	
4. 31-40	10. 91-100	16. 151-160	22. 211-220	28. 271-280	
5. 41-50	11. 101-110	17. 161-170	23. 221-230	29. 281-290	
6. 51-60	12. 111-120	18. 171-180	24. 231-240	30. 291-300	

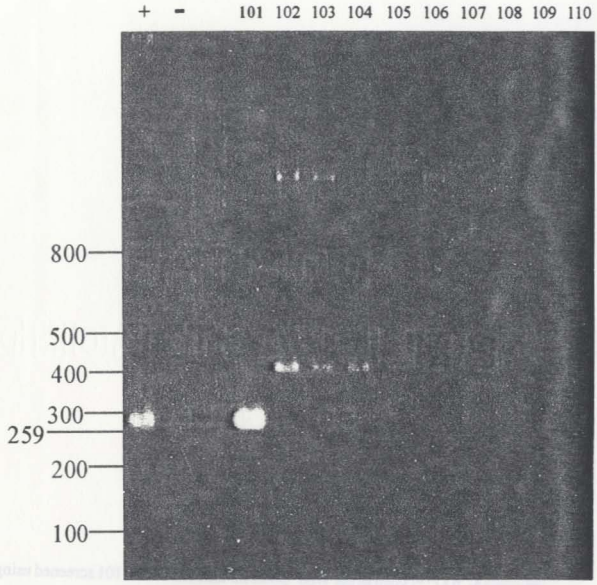
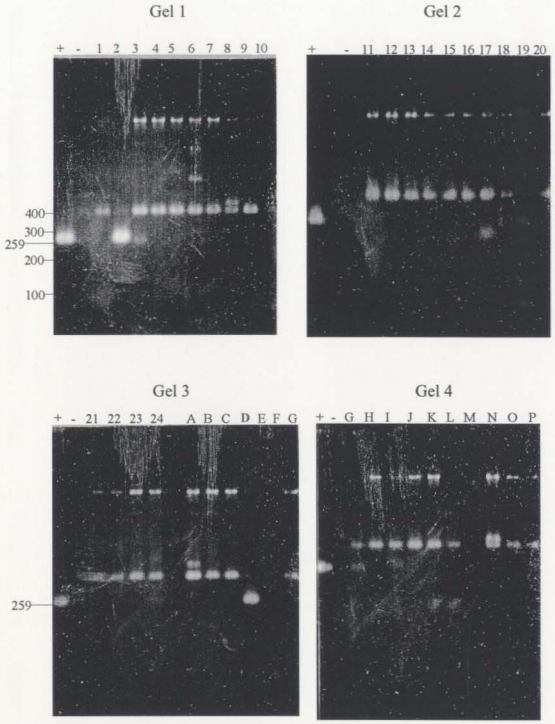


Figure 11: Plate pool screening of the PAC "Plate Pools" with the marker 855R. Note that the marker amplified a fragment in 101, indicating that the PAC of interest is in that plate. + = positive control (human genomic DNA), - = negative control (dH_2O).

Figure 12: PAC "Down to the Well" pools for the PAC in dish 101 screened using the 855R marker. Note that the PAC of interest is in column D and row 2, therefore the name of the PAC is 101d2. + = Positive control (human genomic DNA), - = Negative control (dH_2O).


Gel 1
+ - 1 2 3 4 5 6 7 8 9 10
400
300
259
200
100
Gel 2
+ - 11 12 13 14 15 16 17 18 19 20
Gel 3
+ - 21 22 23 24 A B C D E F G
259
Gel 4
+ - G H I J K L M N O P

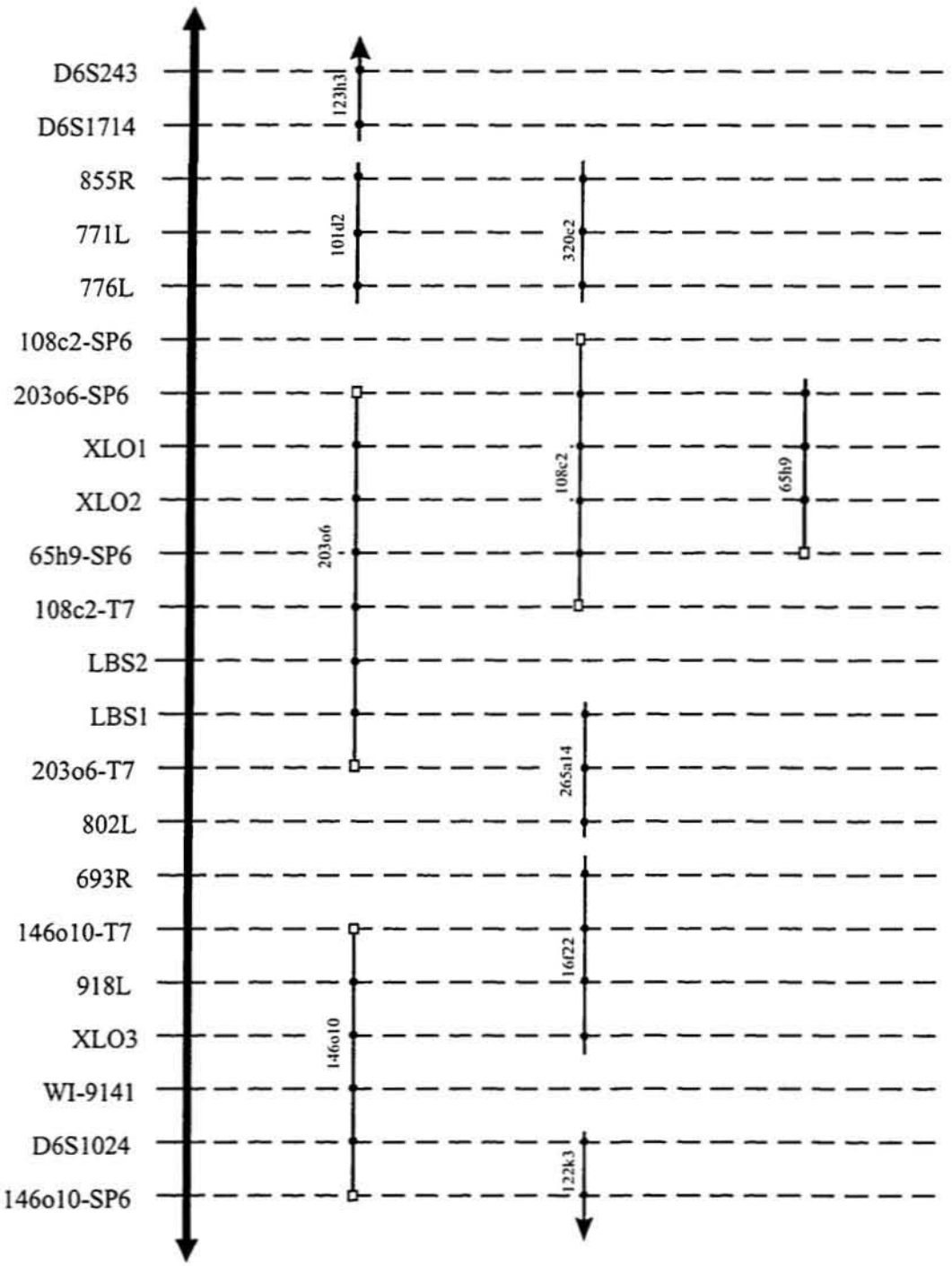
PAC library in an attempt to attain a complete contiguous map of the region (Table 7). Three new PAC clones were isolated with the second screening of the PAC pools giving a total of 10 PAC clones within the ARPKD region. The ends of the new clones were sequenced with the T7 or SP6 primers and the sequences were used to design primers for STS markers. These markers were subsequently used to orient the PAC clones and determine the overlap with other PACs (Fig. 13). Unfortunately, after the second round of PAC clone isolation a complete contiguous map had still not been obtained. At this point it was decided that, due to time constraints, there would not be a third screen of the PAC pool library but rather subcloning of the available PACs should begin.

Table 7: Results of second screen of the PAC library using novel STSs generated from PAC ends.

PAC STS	Annealing temperature (°C)	# of possible PACs Initially Found	Final # of New PACs Found Accurately	PACs Identified
203o6-SP6	55	3	2	108c2, 65h9
146o10-T7	52	3	1	16f22
320c12-T7*	55	0	0	N/A
320c12-SP6*	58	0	0	N/A

* Due to high background annealing no clones were able to be isolated.

Figure 13: Partially contiguous PAC map of the ARPKD region.


D6S243
D6S1714
855R
771L
776L
108c2-SP6
203o6-SP6
XLO1
XLO2
65h9-SP6
108c2-T7
LBS2
LBS1
203o6-T7
802L
693R
146o10-T7
918L
XLO3
WI-9141
D6S1024
146o10-SP6
123h3
101d2
320c2
108c2
65h9
203o6
265a14
16f22
146o10
122k3

5. SEARCHING FOR NOVEL MICROSATELLITE MARKERS WITHIN THE ARPKD REGION

5.1 SCREENING SUBCLONES FOR MICROSATELLITES

Each PAC clone was subcloned into the plasmid vector pUC18 and the subclones were screened with repetitive di- or tri- nucleotides. Seven colonies appeared to hybridize to one of the three probes (Table 8). For those clones which were potentially positive, the colony was picked and the plasmid DNA was isolated. The insert was digested out and a Southern blot was performed using the same probes (Fig. 14).

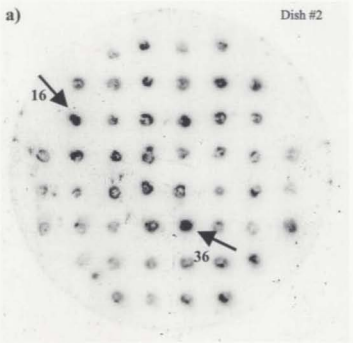
5.2 ANALYZING THE INSERT SEQUENCES

Clones which gave strong positive results after Southern blot analysis were sent to be sequenced (Appendix 3) in order to identify the nature of the repeat and its surrounding sequence (Appendix 7). Only four repeat sequences were found in the subclones one of which was considered too small to be polymorphic (16f22 2-22). It is likely that other repeats were present but were not found as the whole insert was too large to sequence completely without creating more oligonucleotide primers and resequencing. Some of the subclones, for example 65h9 1-9 (F) and 265a14 1-3(R), were found to contain multimeric

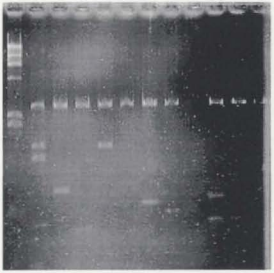
Table 8: Number of subclones produced from each PAC and the number suspected of containing satellite repeats.

PAC clone	# subclones screened for satellite repeat	# suspected satellite repeats
320c12	71	0
101d2	136	0
320c12	110	0
203o6	464	2
265a14	173	1
146o10	416	0
108c2	349	0
65h9	475	2
16f22	497	2

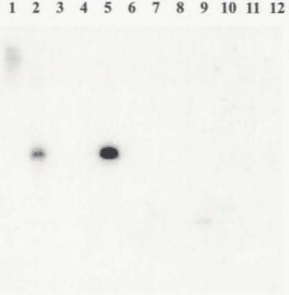
Figure 14: a) Dish #2 of the 203o6 subclone colony screens. Arrows depict two possible inserts containing repeat regions which were named 203o6 2-16 and 2-36. b) PAC 203o6 subclones with the insert fragment digested out. c) Autoradiograph of the Southern hybridization of the gel depicted in b). Note that subclones in lane 2 and 5 have given intense signals indicating the strong possibility that there is a repeat region in each of these clones. d) Table detailing the contents of each lane in the gel in b) and the Southern blot in c). Note that the subclone 203o6 2-36 depicted in a) gave a strong signal in the Southern blot in c).



b)



c)



d)

Lane	Contents	Lane	Contents	Lane	Contents	Lane	Contents
1	λ-HindIII	4	203o6 2-16	7	203o6 3-5	10	203o6 4-26
2	203o6 1-2	5	203o6 2-36	8	203o6 3-17	11	203o6 4-40
3	203o6 1-45	6	203o6 3-1	9	203o6 4-22	12	203o6 9-46

inserts which occurred as a result of the digested PAC fragments reannealing at their Sau3A sites. Oligonucleotide primers were designed around the repeat sequence and used to determine whether or not the repeat was polymorphic in the U1 family(Appendix 8). Unfortunately, when the primers were being designed an error was made and the 5' end of the reverse primer for 16f22 1-40 extends into the pUC18 vector by 5bp. This error also occurred in the reverse primer of 146o10 5-37 but only by 1bp. Two of the satellite repeats which were discovered were monomorphic in the U1 family (16f22 1-40 and 146o10 5-37) and the third, 146o10 5-37, did not work despite several attempts using different reaction conditions. The sequenced insert regions were inputted into NCBI's BLASTN search in order to determine if there were any previously classified sequences (Appendix 9). During the BLAST search one of the 65h9 subclones was found to be completely overlap a known sequence which turned out to be the 108c2 PAC clone which had been completely sequenced, by the Sanger Centre (Hinxton, Cambridge), and entered into NCBI's database 10 days earlier on September 28, 1999 (GenBank accession # AL034343).

6. ANALYZING THE 108C2 PAC SEQUENCE

6.1 ANALYSIS AND MANIPULATION

The sequence of the 108c2 clone was preliminarily analyzed by using the "find and replace" function of WordPerfect to look for simple repeat sequences. Later the program "Sputnik" was obtained and used.. The highest score which resulted from the 108c2 sequence was 59 found at position 14662 - 14736 and resulted from a CTTT tetranucleotide repeat 75 bp in length (Appendix 10). Closer inspection of this region revealed a more complex repeat consisting of di- tri- and tetranucleotide repeats consisting of CT nucleotides with a total size of 289bp. Oligonucleotide primers were designed around this region in order to determine whether or not the marker was variable (Table 9).

Table 9: Properties of the oligos designed around the 108c2 repeat (from "Primer 3").

Marker	Sequence	Tm* (°C)	GC%†	any‡	3'a	annealing temp. (°C)
108c2-FULL1	(F) GCAGAAGGCAGGACAGGTAT	59.31	55.00	2.0	2.0	55
	(R) ATGGAGTGACACCCCGTCT	60.39	57.89	7.0	3.0	

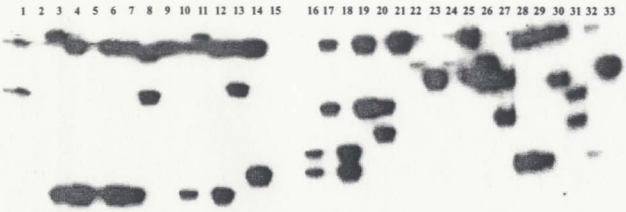
* Melting Temperature
† GC content of oligonucleotide
‡ The self-complementarity score of the oligo as a measure of its tendency to anneal to itself or form secondary structure.
a The self-complementarity of the oligo as a measure of its tendency to form a primer-dimer with itself.

6.2 GENOTYPING

6.2.1 Results and Discussion

The 108c2 FULL1 marker was used to screen the DNA from the U1 family and it was indeed quite polymorphic resulting in a total of 11 different alleles (Fig. 15). The gel was scored and the genotypes were added to the haplotypes previously found in the ARPKD region on chromosome 6 it was then clear that the affected individuals were heterozygous for this marker (Fig. 16). The affected individuals in sibship A all inherited the genotype 2, 20 while those in sibship B inherited the genotype 2, 10 which differed from their unaffected sibling who inherited an 18, 16 genotype. The affected individual in sibship C inherited a genotype of 2, 7 while his sister inherited a 5, 7 genotype. This indicated that the meiotic recombination which occurred in the father of sibship A (PID29) had occurred further towards the telomeric end of chromosome 6, to at least the 108c2 FULL1 marker. Linkage analysis was then carried out in order to determine the statistical significance of this finding (Table 10). Because it is believed that sibship C does not have the same disease as sibships A and B, it was thought to be prudent to determine the linkage of 108c2 FULL1 without the influence of the data from sibship C. Table 10 does show that the LOD scores were higher when considering only sibships A and B; however, at $\theta = 0.00$ the LOD score was still well below 3.0 without the data from sibship C. The two point linkage analysis reflect the hypothesis that due to the recombination in individual PID 29, described in section 3.4.1, the 108c2 FULL1 marker can be excluded from the PKHD-1 critical region.

a)



b)

Lane	PID	Type	Lane	PID	Type	Lane	PID	Type	Lane	PID	Type
1	10	2, 8	10	7	2, 20	19	54	2, 10	28	73	2, 18
2	30	2, 18	11	85	1, 2	20	22	10, 12	29	74	2, 18
3	29	1, 20	12	51	2, 20	21	24	2, 2	30	10	2, 8
4	49	2, 20	13	75	2, 7	22	38	2, 5	31	22	10, 12
5	50	2, 20	14	40	2, 18	23	65	7, 7	32	30	2, 18
6	49	2, 20	15	41	10,16	24	38	2, 5	33	65	7, 7
7	51	2, 20	16	52	16, 18	25	55	2, 7			
8	103	3, 8	17	53	2, 10	26	112	5, 7			
9	48	2, 2	18	52	16, 18	27	78	5, 7			

Figure 15: a) Autoradiograph of U1 family DNA analyzed with the 108c2 FULL1 marker. b) Identification of the DNA samples in each lane of the autoradiograph in a) and the corresponding genotype.

Figure 16: Core pedigree of the U1 family with all affected individuals, their parents, grandparents and other select relatives depicting their haplotypes for the markers in the ARPKD region of chromosome 6 including the new marker 108c2 FULL-1. Note that the meiotic recombination which in individual PID29 occurred above (telomeric) to the 108c2 FULL-1 marker. [ ] = inferred genotype, - = unknown genotype, x = meiotic recombination.

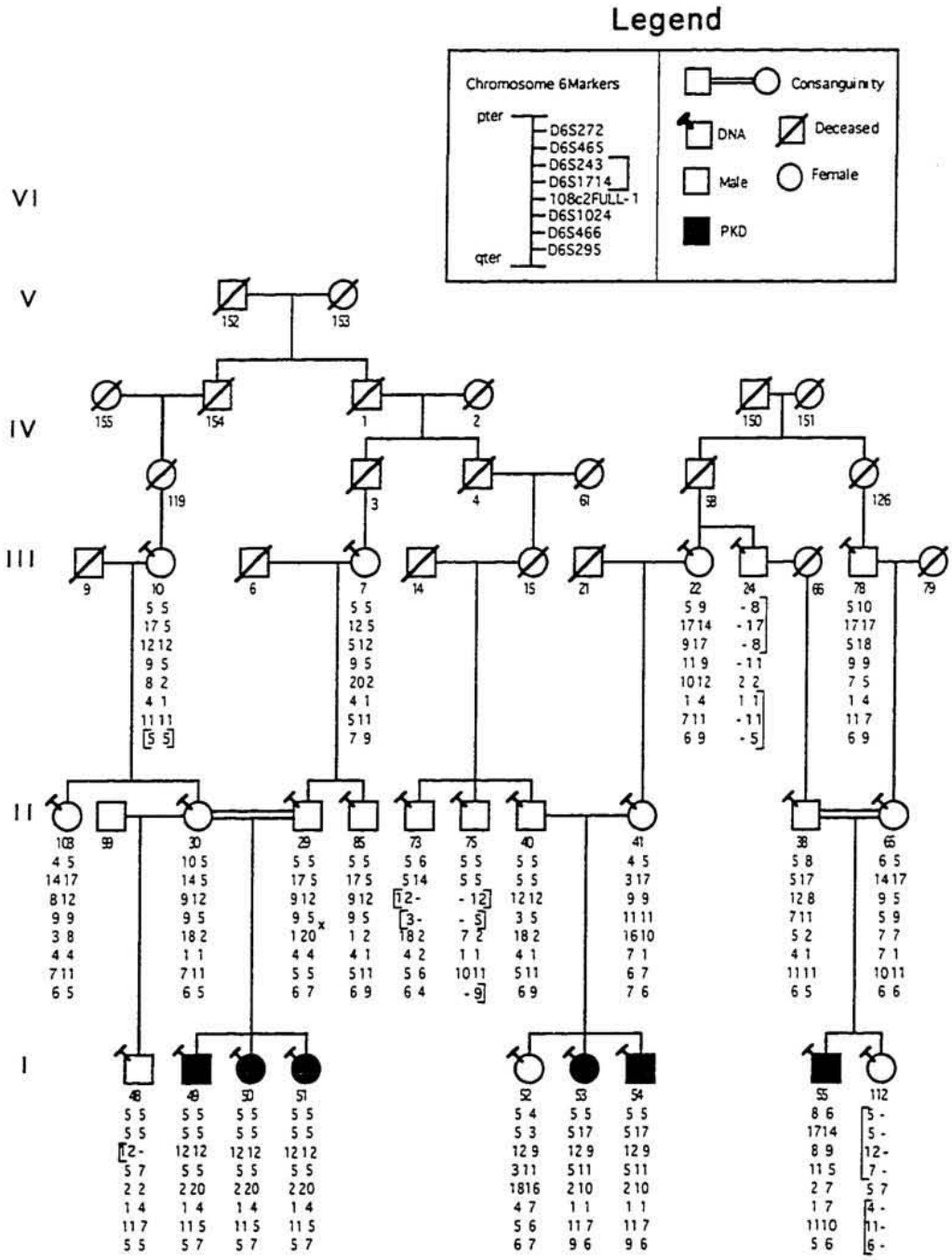

Legend
Chromosome 6Markers
pter
D6S272
D6S465
D6S243
D6S1714
108c2FULL-1
D6S1024
D6S466
D6S295
qter
Consanguinity
DNA
Deceased
Male
Female
PKD
VI
V
IV
III
II
I

Table 10: LOD scores of two-point linkage analysis on the novel marker in the ARPKD region.

Marker	$\theta = 0.00$	$\theta = 0.05$	$\theta = 0.10$
108c2 FULL1 (Sibships A, B and C)	0.05	1.77	2.02
108c2 FULL1 (Sibships A and B only)	1.48	2.23	2.11

Using a contiguous map recently produced by Park et al. (1999) it was possible to estimate the size of the region between 108c2 FULL1 and D6s1714/D6s243 which contains the PKHD-1 gene. Through knowledge of the size of each of the PAC or BAC clones found within this region it was possible to estimate the distance between each of the markers. By doing so it was estimated that a region of 516Kb between markers 108C2FULL1 and D6S1024 could be ruled out and the critical region now lies between the108C2FULL1 and D6S1714/D6S243 markers, a region of approximately 560Kb in size. This finding was significant as it essentially cut the size of the critical region in half leaving a relatively small region to be screened for expressed sequences. It is estimated there are 20-25 genes per Mb (Fields et al., 1994). Therefore it can be inferred that only 11 - 14 genes remain to be screened for mutation in individuals affected by ARPKD.

7. CONCLUSION

Homozygosity mapping with polymorphic markers for several PKD loci made it possible to determine if the affected individuals in sibship A were HBD in a particular disease region. It was also possible to determine that even though the individuals in sibship B were not HBD, it is likely that they have inherited two different mutations of the PKHD1 gene. Other individuals with different phenotypic variants of ARPKD in Newfoundland are known to clinicians. Therefore it is quite possible for affected individuals who are not HBD to inherit different mutations of PKHD1 (Dr. Jane Green, personal communication). LOD score analysis of the resulting haplotypes enabled the elimination of three potential diseases; NPH-1 and -2 on chromosomes 2 and 10, respectively, and a disease syntenic to the mouse bpk/jcpk gene on human chromosome 10 (Fig. 5, 6 and 7). The kidney pathology in sibships A and B was linked to an area of chromosome 6 which is known to be the location of the ARPKD gene, PKHD-1, and suggested the presence of two separate mutations. Affected individuals in sibship A were homozygous for one haplotype and the affected individuals in sibship B each had one copy of this haplotype and a copy of another haplotype. Initial analysis of the region using known microsatellite markers revealed a meiotic recombination on PID 29 which was inherited by all three of his affected children. This enabled the narrowing of the centromeric end of the ARPKD critical region to a 1cM region flanked centromerically by the marker D6S1024. The region was previously determined to be flanked by the markers D6S243/D6S1714 on the telomeric end of the region by Guay-

Woodford et al. (1995) (Fig. 8). The affected individual in sibship C (PID 55) did not share either of the common haplotypes seen in the affected individuals of sibship A and B. This confirmed the clinical data which indicated that he had a different disease, possibly ADPKD. Kidney ultrasound failed to reveal any significant cysts in the parents of PID 55, indicating that ADPKD was not inherited by this individual from his parents. However, this does not rule out the possibility of a novel mutation of the ADPKD 1 or 2 gene. This will only be confirmed once this individual can be screened for a mutation in each of the two ADPKD genes.

Upon production of a PAC map and screening PACs for microsatellites, new repeat regions were found within the critical region but they were not variable in the U1 family. The elucidation of the complete sequence of the 108c2 clone by the Sanger Centre and its publication to the public domain led to the discovery of the novel repeat sequence 108c2 FULL-1 approximately 14, 000bp from the telomeric (SP6) end of the clone. Analysis of the U1 family with this new marker revealed that it was highly variable and informative. The analysis revealed that the recombination in PID 29 which allowed the narrowing of the ARPKD critical region to a 1cM area had occurred further towards the telomeric end of the chromosome, producing a new critical region of approximately 560kb.

Future research in this area will focus on finding expressed sequences in the critical region using the techniques of exon trapping and kidney cDNA library screening. Screening can be done using the remaining PACs in the critical region, 320c12, 101d2, 123h3 and others named in Park et al. (1998) as probes. There are genes in the region which have

already been ruled out as the PKHD-1 gene. They include; the alpha subunit of the metalloendopeptidase meprin (Jiang et al., 1995), gluthathione S-transferase (Board and Webb, 1987), methylmalonyl coenzyme A mutase (Dib et al., 1996) (reviewed by Zerres et al., 1998), retinal degeneration, slow and casein kinase II, β subunit (Mücher et al., 1998) and most recently, the KIAA0057 gene which encodes a TRAM-like protein (Onuchic et al., 1999).

Given the relatively small critical region for ARPKD and the number of investigators working on finding the PKHD-1 gene, it is entirely likely that the gene which causes the disease will be found within the next 6 to 12 months. In fact, there has already begun an intensive search by an international consortium of laboratories who appear to be quite close to finding the PKHD-1 gene. Once this task has been accomplished, research will then focus on determining the structure and function of the gene product and finding means to offset the effects of the disease.

REFERENCES

Akhilesh, P. and Lewitter, F. (1999). Nucleotide sequence databases: a gold mine for biologists. *Trends in Biochemical Science.* **24**: 276-280

Antignac, C., Arduy, C., Beckmann, J. S., Benessy, F., Gros, F., Medhioub, M., Hildebrandt, F., Dufier, J. L., Kleinknecht, C., Broyer, M., Wessenbach, J., Habib, R. and Cohen, D. (1993). A gene for familial Juvenile nephronophthisis (recessive medullary cystic disease) maps to chromosome 2p. *Nature Genetics.* **3**: 342-345.

Avner, E. D. (1995). Role of growth factors in cell proliferation. *Kidney International.* **47**: 719-720.

Avner, E. D. and Sweeney, W. E. (1995). Apical and epidermal growth factor receptor expression defines a distinct cystic tubular epithelial phenotype in autosomal recessive polycystic kidney disease. *Pediatric Research.* **37**: 359

Board, R. G. and Webb, G. C. (1987). Isolation of a cDNA clone and localization of human glutathionine S-transferase 2 genes to chromosome band 6p12. *Proceedings National Academy Science U. S. A..* **84**: 2377-2381.

Burke, D. T., Carle, G. F. and Olson, M..V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science*. **236**: 806-12.

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

Bryda, E. C., Ling, H., Rathbun, D. E., Burmeister, M. and Flaherty, L. (1996). Fine genetic map of mouse chromosome 10 around the polycystic kidney disease gene, jcpk, and Ankyrin3. *Genomics*. **35**: 425-430.

Calvet, J. P. (1998). Molecular genetics of polycystic kidney disease. *Journal of Nephrology*. **11**: 24-34.

Carrano, A. V., de Jong, P. J., Branscomb, E., Slezak, T., Watkins, B. W. (1989). Constructing chromosome- and region-specific cosmid maps of the human genome. *Genome*. **31**: 1059-1065.

Chauveau, D., Pirson, Y., Le Moine, A., Franco, D., Belghiti, J., Grunfeld, J. P. (1997). Extrarenal manifestations in autosomal dominant polycystic kidney disease. *Advances in Nephrology*. **26**: 265 - 289.

Collins, F. S. (1992a). Identifying human disease genes by positional cloning. *The Harvey Lectures Series*. **86**: 149-164.

Collins, F. S. (1992b). Positional cloning: let's not call it reverse anymore. *Nature Genetics*.**1**: 3-6.

Cottingham, R. W., Jr., Idury, R. M., and Schäffer, A. A. (1993). Faster sequential genetic linkage computations. *American Journal of Human Genetics*. **53**: 252-263.

Deget, F., Rudnik-Schöneborn and Zerres, K. (1995). Course of autosomal recessive polycystic kidney disease (ARPKD) in siblings: a clinical comparison of 20 sibships. *Clinical Genetics*. **47**: 248-253.

Dib, C., Fauré, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissett,e J., Weissenbach, J. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**:152-154.

Duyk, G. M., Kim, S., Myers, R., Cox, D. R. (1990). Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proceedings of the National Academy Science U. S. A.* **87**: 8995-8999.

Fanconi, G., Hanhart, E., von Albertini, A., Euhlmger, R., Dolive, G., and Prader, A. Die familiare juvenile Nephronophthier. *Helvetica Paediatrica-Acta.* **6**: 1-9.

Feingold, J. M., Ogden, S. D., Denny, C. T. (1990). Streamlined approach to creating yeast artificial chromosome libraries from specialized cell sources. *Proceedings National Academy Science U. S. A.*. **87**(21): 8637-8641.

Fick, G. M. and Gabow, G. A. (1994). Hereditary and acquired cystic disease of the kidney. *Kidney International.* **46**: 951-964.

Fields C., Adams M. D., White O., Venter J. C. (1994). How many genes in the human genome? *Nature Genetics.* **7**: 345-346.

Gang, D. L. and Herrin, J. T. (1986). Infantile polycystic disease of the liver and kidneys. *Clinical Nephrology.* **25**: 28-36.

Gillessen-Kaesbach, G., Meinecke, P., Garrett, C., Padberg, B. C., Rehder, H. and Passarge, E. (1993). New autosomal recessive lethal disorder with polycystic kidneys type potter I, characteristic face, microcephaly, brachymelia, and congenital heart defects. *American Journal of Medical Genetics.* **45**: 511-518.

Gattone, V. H., Andrews, G. K., Niu, F. W., Chadwick, L. J., Klein, R. M. and Calvet, J. P. (1990). Defective epidermal growth factor gene expression in mice with polycystic kidney disease. *Developmental Biology*. **138**: 225-230

Guay-Woodford, L., Müecher, G., Hopkins, D., Avner, E. D., Germino, G. G., Guilliot, A. P., Herrin, J., Holleman, R., Irons, D. A., Primack, W., Thomson, P. D., Waldo, F. B., Lunt, P. W. and Zerres, K. (1995). The severe perinatal form of autosomal recessive polycystic kidney disease maps to chromosome 6p21.1-p12: Implications for genetic counseling. *American Journal of Human Genetics*. **56**: 1101-1107.

Guay-Woodford, L. Bryda, E. C., Christine, B., Lindsey, J. R., Collier, W. R., Avner, E. D., D'Eustachio and Lorraine Flaherty (1996). Evidence that two phenotypically distinct mouse PKD mutations, *bpk* and *jcpk* are allelic. *Kidney International*. **50**; 1158-1165.

Gusmano, R., Ghiggeri, G. M. and Caridi, G. (1998). Nephronophthisis - medullary cystic disease: clinical and genetic aspects. *Journal of Nephrology*. **5**: 224-228.

Haider, B., Carmi, R., Shalev, H., Sheffield, V. C. and Landau, D. (1998). A Bedouin kindred with infantile nephronophthisis demonstrates linkage to chromosome 9 by homozygosity mapping. *American Journal of Human Genetics*. **63**: 1404-1410.

Hildebrandt, F. (1998). Identification of a gene for nephronophthisis. *Nephrology Dialysis Transplantation*. **13**: 1334-1336.

Iakoubova, O. A., Dushkin, H. and Beier, D. R. (1995). Localization of a murine recessive polycystic kidney disease mutation and modifying loci that affect disease severity. *Genomics*. **26**: 107-114.

Iglesias, D. M., Telleria, D., Viribay, M., Herrera, M., Bernath, V. A., Kornblihtt, A. R., Martin, R. S. and Millan, J. L. (2000). A novel frameshift mutation (2436insT) produces an immediate stop codon in the autosomal dominant polycystic kidney disease 2 (PKD2) gene. *Nephrology Dialysis Transplantation*. **15**: 477-480.

Ioannou, P. A., Amemiya, C. T., Garnes, J., Kroisel, P. M., Shizuya, H., Chen, C., Batzer, M. A., de Jong, P. J. (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genetics*. **6**: 84-89

Jiang, W., Dewald, G., Brundage, E., Mucher, G., Schildhaus, H. U., Zerres, K. and Bond, J. S. (1995). Fine mapping of MEP1A, the gene encoding the alpha subunit of the metalloendopeptidase meprin, to human chromosome 6p21. *Biochemical and Biophysical Research Communications*. **216**:630-635.

Jones, G. (1997). Identification of autosomal recessive polycystic kidney disease in a Newfoundland family. B.Sc. Honours thesis, Memorial University of Newfoundland.

Kim, U. K., Jin, D. K., Ahn, C., Shin, J. H., Lee, K. B., Kim, S. H., Chae, J. J., Hwang, D. Y., Lee, J. G., Namkoong, Y. and Lee, C. C. (2000). Novel mutations of the PKD1 gene in Korean patients with autosomal dominant polycystic kidney disease. *Mutation Research*. **432**: 39-45.

Konrad, M., Saunier, S., Calado, J., Gubler, M., Broyer, M. and Antignac, C. (1998). Familial juvenile Nephronophthisis. *Journal of Molecular Medicine*. **76**: 310-316.

Lander, E. and Botstein, D. (1987). Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science*. **236**: 1567-1570.

Lathrop, G. M., and Lalouel, J. M. (1984). Easy calculations of LOD scores and genetic risk on small computers. *American Journal of Human Genetics*. **36**: 460-465.

Lens, X. M., Onuchic, L. F., Wu, G., Hayashi, T., Daoust, M., Mochizuki, T., Santarina, L. B., Stockwin, J. M., Mücher, G., Becker, J., Sweeny Jr., W. E., Avner, E. D., Guay-Woodford, L., Zerres, K., Somlo, S. and Germino, G. G. (1997). An integrated

genetic and physical map of the autosomal recessive polycystic kidney disease region. *Genomics*. **41**: 463-466.

Leppert, M. F. (1990). Gene mapping and other tools for discovery. *Epilepsia*. **31**: Suppl 3:S11-8.

Lovett, M., Kere, J. and Hinton, L. (1991). Direct Selection: A method for the isolation of cDNAs encoded by large genomic regions. *Proceedings of the National Academy of Science USA*. **88**: 9628-9632.

Medhioub, M., Cherif, D., Benessy, F., Silbermann, F., Gubler, M. C., Le Paslier, D., Cohen, D., Weissenbac, J., Beckmann, J. and Antignac, C. (1994). Refined mapping of a gene (NPH1) causing familial nephronophthisis and evidence for genetic heterogeneity. *Genomics*. **22**: 296-301.

Merta, M, Štekrová, J., Židovská, M., Voršilkova, J., Walterová, J., Kapras, J., Alánová, M., Rašková, D., Gaillyová, R., Archmanová, E. and Ryšavá, R. (1997). DNA diagnosis and clinical manifestations of autosomal dominant polycystic kidney disease. *Folia Biologica (Paraha)*. **43**: 201-204.

Miller, S. A., Dykes, D. D., Polesky, H. F. (1988). A simple salting out procedure for

extracting DNA from human nucleated cells. *Nucleic Acids Research*. **16**: 1215.

Morton, N. E. (1955). Sequential tests for the detection of linkage. *American Journal of Human Genetics*. **7**: 277-318.

Moyer, J. H., Lee-Tischler, M. J., Kwon, M. J., Schrick, J. J., Avner, E. D., Sweeney, W. E., Godfrey, V. L., Cacherio, N. L. A., Wilkinson, J. E., Woychik, R. P. (1994). Candidate gene associated with mutation causing recessive polycystic kidney disease in mice. *Science*. **264**: 1329-1333.

Mücher, G., Becker, J., Knapp, M., Büttner, R., Moser, M., Rudnik-Schöneborn, S., Somlo, S., Germino, G., Onuchic, L., Avner, E., Guay-Woodford, L. and Zerres, K. (1998). Fine mapping of the autosomel recessive polycystic kidney disease locus (PKHD1) and the genes MUT, RDS, CSNK2β and GSTA1 at 6p21.1-p12. *Genomics*. **48**: 40-45.

Mücher, G., Brunhilde, W. and Zerres, K. (1994). Refining the map and defining flanking markers of the gene for autosomal recessive polycystic disease on chromosome 6p21.1-p12. *American Journal of Human Genetics*. **55**: 1281-1284.

Neuhausen, S. L., Swensen, J., Miki, Y., Liu, Q., Tavtigian, S., Shattuck-Eidens, D., Kamb,

A., Hobbs, M. R., Gingrich, J., Shizuya, H., Kim, U., Cochran, C., Futreal, P., Wiseman, R., Lynch, H., Tonin, P., Narod, S., Cannon-Albright, L., Skolnick, M., Goldgar, D. (1994). A P1-based physical map of the region from D17S776 to D17S78 containing the breast cancer susceptibility gene BRCA1. *Human Molecular Genetics*. **3**(11): 1919-1926.

Ogborn, M. R. (1994). Polycystic kidney disease - a truly pediatric problem. *Pediatric Nephrology*. **8**: 762-767.

Onuchic L. F., Mrug M., Lakings A. L., Muecher G., Becker J., Zerres K., Avner E. D., Dixit M., Somlo S., Germino G. G., Guay-Woodford L. M. (1999). Genomic organization of the KIAA0057 gene that encodes a TRAM-like protein and its exclusion as a polycystic kidney and hepatic disease 1 (PKHD1) candidate gene. *Mammalian Genome*. **10**: 1175-8

Orellana, S. A., Sweeney, W. E., Neff, C. D. and Avner, E. D. (1995). Epidermal growth factor receptor expression is abnormal in murine polycystic kidney. *Kidney International*. **47**: 490-499.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation

polymorphisms. *Proceedings of the National Academy of Science U S A.* **86**: 2766-2770.

Papadopoulos, N., (1995). Introduction to positional cloning. *Clinical and Experimental Allergy.* **25**; 116-118.

Park, J. H., Dixit, M. P., Onuchic, L. F., Wu, G., Goncharuk, A. N., Kneitz, S., Santarina, L. B., Hayashi, T., Avner, E. D., Guay-Woodford, L., Zerres, K., Germino, G. G. and Somlo, S. (1999). A 1-Mb BAC/PAC-based physical map of the autosomal recessive polycystic kidney disease gene (PKHD1) region on chromosome 6. *Genomics.* **57**: 249-255.

Proesmans, W., Van Damme, B., Macken, J. (1974). Nephronophthisis and tapetoretinal degeneration associated with liver fibrosis. *Clinical Nephrology.* **3**:160 - 164.

Richards, W. G., Sweeney, W. E., Yoder, B. K., Wilkinson, J. E., Woychik, R. P. and Avner, E. D. (1998). Epidermal growth factor receptor activity mediates renal cyst formation in polycystic kidney disease. *Journal of Clinical Investigation.* **101**: 935-939.

Robson, K. J., Chandra, T., MacGillivray, R. T. and Woo, S. L. (1982). Polysome

immunoprecipitation of phenylalanine hydroxylase mRNA from rat liver and cloning of its cDNA. *Proceedings of the National Academy of Science, U. S. A.* **79**: 4701-4705.

Schäffer, A. A, Gupta, S. K., Shriram, K. and Cottingham, R. W., Jr. (1994). Avoiding recomputation in linkage analysis. *Human Heredity.* **44**: 225-237.

Sessa, A., Ghiggeri, G. M. and Turco, A. E. (1997). Autosomal dominant polycystic kidney disease: clinical and genetic aspects. *Journal of Nephrology.* **10**: 295-310.

Sheffield, V., Stone, E. M. and Carmi, R. (1998). Use of isolated inbred human populations for identification of disease genes. *Trends in Genetics.* **14**: 391-396.

Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y., Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proceedings of the National Academy Science U.S.A.*. **89**: 8794-8797.

Smith, C. H., and Graham, Y. B. (1945). Congenital medullary cystic disease of the kidney with severe refractory anemia. *American Journal of Diseases of Children.* **69**: 369-377.

Sternberg, N. L., (1992). Cloning high molecular weight DNA fragments by the bacteriophage P1 System. *Trends in Genetics*. **8**: 11-16.

Strachan, T. and Read, A. P. Human Molecular Genetics 2, John Wiley and Sons Inc., 1999. Ontario.

Sweeney, W. E. avner and E. D. (1998). Functional activity of epidermal growth factor receptors in autosomal recessive polycystic kidney disease. *American Journal of Physiology*. **275**: 387-394.

The Utah Marker Development Group (1995). A Collection of ordered tetranucleotide-repeat markers from the human genome. *American Journal of Human Genetics*. **57**: 619-628.

Thompson, M. W., McInnes, R. R. and Willard, H. F. Genetics in Medicine, W. B. Saunders Company, 1991. Philadelphia. pp. 183-184, 435.

Traver, C. N., Klapholz, S., Hyman, R. W. and Davis, R. W. (1989). Rapid screening of a human genomic library in yeast artificial chromosomes for single-copy sequences. *Proceedings of the National Academy of Science U S A*. **86**: 5898-902.

Tsui, L. C., Buchwald, M., Barker, D., Braman, J. C., Knowlton, R., Schumm, J. W., Eiberg, H., Mohr, J., Kennedy, D., Plavsic, N., et al. (1985). Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science*. **230**: 1054-1057.

Uberbacher, E. C. and Mural, R. J. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proceedings of the National Academy of Science U S A*. **88**: 11261-11265.

Watson, J. D., Gilman, M., Witkowski, J. and Zoller, M. Recombinant DNA, Scientific American Books 1992, New York. pp. 522.

Whittaker, P. A., Campbell, A. J., Southern, E. M., Murray, N. E. (1988). Enhanced recovery and restriction mapping of DNA fragments cloned in a new lambda vector. *Nucleic Acids Research*. **16**: 6725-36.

Wilson, P. D., Du, J., and Norman, J. T. (1993). Autocrine, endocrine and paracrine regulation of growth abnormalities in autosomal dominant polycystic kidney disease. *European Journal of Cell Biology*. **61**: 131-138.

Witzleben, C. L. and Sharp, A. R. (1982). "Nephronophthisis-congenital hepatic fibrosis": an additional hepatorenal disorder. *Human Pathology*. **13**: 728 - 733.

Young, T. L., Penney, L., Woods, M. O., Parfrey, P. S., Green, J. S., Hefferton, D. and Davidson, W. S. (1999). A fifth locus for Bardet-Biedl syndrome maps to chromosome 2q31. *American Journal of Human Genetics*. **64**: 900-904.

Zerres, K., Hansmann, M., Mallmann, R. and Gembruch, U. (1988) Course of autosomal recessive polycystic kidney disease (ARPKD) in siblings: a clinical comparison of 20 sibships. *Prenatal Diagnosis*. **8**: 215-229

Zerres, K., Kohlschmidt, N, Mücher, G., Hallermann, C., Shahidi, P., Welleck, B., Theile, U., Schumacher, R., Bahlmann, F., Rudnik, S., Muentefering, H. (1999). A syndrome with autosomal recessive polycystic kidney with skeletal and facial anomalies is not linked to the ARPKD gene locus on chromosome 6p. *American Journal of Human Genetics*. **65**: 1972.

Zerres, K., Mücher, G., Bachner, L., Deschennes, G., Eggermann, T., Kääriäinen, H., Knapp, M., Lennert, T., Misselwitz, J., von Mühlendahl, K. E., Neumann, H. P. H.., Pirson, Y. Rudnik-Schöneborn, S., Steinbricker, V., Wirth, B. and Schärer, K. (1994). Mapping of the gene for autosomal recessive polycystic kidney disease (ARPKD) to chromosome 6p21-cen. *Nature Genetics*. **7**: 429-432.

Zerres, K., Mücher, G., Becker, J., Steinkamm, C., Rudnik-Schöneborn, S., Heikkilä, P.,

Rapola, J., Salonen, R., Germino, G. G., Onuchic, L., Somlo, S., Avner, E. D., Harman, L. A., Stockwin, J. M. and Guay-Woodford, L. M. (1998). Prenatal diagnosis of Autosomal Recessive Polycystic Kidney Disease (ARPKD): Molecular genetics, Clinical experience, and fetal morphology. *American Journal of Medical Genetics*. **76**: 137-144.

Zerres, K., Rudnik-Schöneborn, S., Deget, F., Holtkamp, U., Brodehl, J., Geisert, J. and Scharer, K. (1996).Course of autosomal recessive polycystic kidney disease (ARPKD) in siblings: a clinical comparison of 20 sibships. *Acta Paediatrica*. **85**; 437-445.

Zerres, K., Rudnik-Schöneborn, S., Steinkamm, C., Becker, J. and Mücher, G. (1998). Autosomal recessive polycystic kidney disease. *Journal of Molecular Medicine*. **76**: 303-309.

Zerres, K., Völpel, M. -C. and Weiß, H. (1984). Cystic kidneys. *Human Genetics*. **68**: 104-135.

APPENDICES

Appendix 1: Characterization of the Microsatellite Markers Used in Haplotype Analysis

Polymorphic markers within the NPH-1 region.

Marker Name[a]	Coordinate (KcM)[*]	Nature of Repeat[†]	Minimum Length (bp)[†]	Maximum Length (bp)[†]	Maximum Heterozygosity[†]
D2S293	140	Dinucleotide	167	191	0.84
D2S340	142	Dinucleotide	149	173	0.71
D2S160	145	Dinucleotide	204	218	0.79
D2S121	146	Dinucleotide	156	184	0.83
D2S363	148	Dinucleotide	254	264	0.69

Polymorphic markers within the NPH-2 region.

Marker Name[a]	Coordinate (KcM)[*]	Nature of Repeat[†]	Min. Length (bp)[†]	Max. Length (bp)[†]	Maximum Heterozygosity[†]	Annealing Temp. (°C)
D9S197	100.8	Dinucleotide	199	215	0.68	55
D9S280	102.0	Dinucleotide	-	-	0.63	55
D9S910	104.5	Trinucleotide	107	-	-	55
D9S938	110.9	Tetranucleotide	370	-	-	55
D9S2026	117.3	Tetranucleotide	127	-	-	55
D9S105	120.0	Dinucleotide	187	203	0.80	55

Polymorphic markers within the bpk/jcpk region.

Marker Name	Coordinate (KcM)*	Nature of Repeat†	Min. Length (bp)†	Max. Length (bp)†	Maximum Heterozygosity†	Annealing Temp. (°C)
D10S196	77	Dinucleotide	101	990	0.79	58
D10S568	79	Dinucleotide	152	-	0.73	55
D10S1756	83	Dinucleotide	-	-	0.84	55
D10S581	90	Dinucleotide	144	-	0.80	55
D10S1670	94	Dinucleotide	-	-	0.76	55

Polymorphic markers within the ARPKD region.

Marker Name[a]	Coordinate (KcM)*	Nature of Repeat†	Min. Length (bp)†	Max. Length (bp)†	Maximum Heterozygosity†
D6S272	75.0	Dinucleotide	180	196	0.73
D6S465	75.0	Dinucleotide	184	198	0.75
D6S243	-	Dinucleotide	170	250	0.72
D6S1714	75.0	Dinucleotide	127	-	0.65
D6S1024	-	Trinucleotide	121	-	-
D6S466	79.0	Dinucleotide	108	119	0.71
D6S295	80.0	Dinucleotide	93	107	0.74

a Bold indicates nearest flanking marker to critical region

* From Généthon (http://gopher.genethon.fr/genethon_en.html)

† From GDB (http://gdbwww.gdb.org/)

- Information not available

Appendix 2: Quantification of DNA at each step of Qiagen PAC DNA isolation

During the process of PAC isolation using Qiagen tips, an experiment was carried out in order to ensure DNA was not being lost throughout the process and to ensure pure DNA was being isolated. Essentially the cell lysate was applied to the calibrated tip and allowed to run through. The tip was washed twice and the PAC DNA was eluted (see materials and methods). A 1mL sample of each wash buffer was taken in order to quantify the DNA which had come off, in addition the eluate was collected in 1mL samples (15mL total).

Sample	Sample concentration (mg/mL/1mL sample)	Total DNA (μg)
lysate flowthrough	0	0
wash 1	17.6	176
wash 2	5.2	52
eluate 1	2.0	2.0
eluate 2	1.6	1.6
eluate 3	0	0
eluate 4	0	0
eluate 5	8.8	8.8
eluate 6	17.2	17.2
eluate 7	6.8	6.8
eluate 8	3.2	3.2

eluate 9	4.4	4.4
eluate 10	2.0	2.0
eluate 11	2.0	2.0
eluate 12	0.8	0.8
eluate 13	1.2	1.2
eluate 14	0.8	0.8
eluate 15	0.8	0.8

These results show that there was no detectable DNA being lost in the flowthrough which is important as it indicated that the tip was not being overloaded. The clear drop in eluted DNA after 9mL led to the decision to elute with 10mL elution buffer which would create a higher concentration of PAC DNA in the eluate and allow for more effective precipitation and a "cleaner" end product.

Appendix 3: Sequencing Primers

Primer Name	Primer Sequence (5' - 3')
T7	CCGCTAATACGACTCACTATAGGG
SP6	GGCCGTCGACATTTAGGTGACAC
Universal	GTAAACGACGGCCAGT
M13 (-48)	AGCGGATAACAATTTCACACAGGA

Appendix 4: PAC end sequences

123h3-T7

CCTGGTTGGAATGGTGATGAACATTTGAGCATGACATCCTGGTTTTTGTACAA

GTTTTCTATTGCTGCTATAACAAATTATTAACAATGTAGTGGCTTCAAACAAC

ACCTAAGTATTATCTCACAGTTTTGTAGGTCAGAAGCTGGTGAACTTGGCTAT

TTTCCACTGGCTATTTTCTCTGCTTCAGATCTCTGATGGCCAAAATTCAGGTA

ATGGCAGGATTGTGTTTCTTTCTGGAGTTCTGGAGATGATTTCACCTTCAGTC

TCACTCAGGTCTGGCCACATCAGTCCATGTAGGTGTAGGACTGTGGTCGTGTA

ACCGCCAAGAGTCCGCTG

123h3-SP6

AGAAGGATCTGTACTCATTCCCTATAATAAAATGCCAAACCAGAAAAGTCAC

AATTTCTTGTGAAATTCAAGTACACTCCTTCTTAAGAAGAGGCTTTAATTGCT

ATTAGATGAGACATGGCAAACAAACTTGTTACCTA...GTTATAACAACTATAA

CATGGCCATGGTCATACATATGTTTAATATATTATCATAGTAGTCNTATATCA

CTGCATATAGCATCTATATGGATGGTCAGAGGTGTGCCCATGAGAGTCGCA

320c12-T7

GATCCACCATCTTATCTTGGTGCCATCCCTGACTTGCCTTCTGTTCATAAACGC

CTATTAAATGTTTCTTTCTGAGAAACTGGATTTGTCAGCCTCTTTCTTTGGGTA

TCTCAGGTTCCTTGGCCTTTGCGGGTAGGTTTATATAGACCTGCTCAGCACAG

GACAGGCAGTTTCTCAAAAAATTGAAAATAGAATTACCAAATGATCCGGCAA

TATCACTTCTGGGTATATAGCCAAAATAATTGAAAGCAAGGTCTCATAGAAA

TATTTGTACACTGATATTTATAGCAGTGGTATTCACACTCGTCAAAAGATGGA

TGCAGCCGAATTGTCCATAGGCAGATGAATTGATAAAATGTGGTATATACAT

ACAGTAAAATATTCTTCAGCCTTAAAAGAAGAAATCTACACATGCTACACAT

TGGATGAACATTGAGACATNTGCTTAGTGAATAGCCNGTTTGAAGACATCCT

GGTCTTNCTTTTTGAAACGTCAACNTGGTAGCACCTATTTGACCAGGNANCCN

GGTTCCCGGCTTNGAAGGGAAATGGAANNTTTTTTNNNANTTNNNNNCTTTN

NANNAAAAATNCCCAANTTTTNTNCNCTTC (GenBank Accession Number AQ53751)

320c12-SP6

GAAGGATCAAACCTCCTTACATAACATCAGGGTGTGTCACAGATGATTGGTG

AGGTGCACTGAGATCAAAGTTAAATTATACTTAAACATCAAACTCTGGGTGC

CTTCAAATTCTTTGATGACTATCATAGGAAAGAGTTACATTCACTAAAATAAT

TAGTCCACTCTTACCTATCTTCAGCCTAACTCAGGATTACACTGAACTGAAAC

CCGCTGGGGTGCATTTTCCAGAAAAGGGCCAAGAAAAGGTAGCAGGAAAAA

TTACCAAGTTGAATGGGAAGGTAATGA(GenBank Accession AQ537510)

101d2-T7

(GenBank Accession Number AQ321186)

101d2-SP6

(GenBank Accession Number AQ321184)

203o6-T7

TCAGATAGCCATCCATACTTAGTTTTTCTATTTAGATCCCCTGCCAAATGGCC

AACTGGGAAAAGTAACCATGATACTAATTTCACATACAGGATGCTAAAGAAC

TGTGGTTAGTCAGATTAAAAATAAAAATCATGATTTTCTTGTACTGTCTGTCT

ATTTTATCCATTTAATGGTGGAATGGTAGGGGCTCACTAAAGCTAATTTGATT

TTCATTAGGAATTGCTCATGCCATTTTTTCCTTTTTGTTTACTTTTAATAATAA

CAATTCAGGGAGCAAATAAATTAACCCATGTAGACAAAAAGAGATGTAGAAAT

(GenBank Accession Number AQ418637)

203o6-SP6

TACTTTGAACCTAGATGACCCACTCCAACGTTCAGCCTGCACGCCCCATCCAA

GCAGTGCCTCCCAGAGCTGATTTCTGTCAAAGCCATGAGAATGTGGTTGTGG

GAAATGGTTACAGGGCTTCTGGTGTTGTCACCACTATAAACAAAGTCTGGTCC

TCACCTCAGGCCATACAAAGGGCTGCGGCGTGTCACCTATAGATAGCACGTA

ACAAAACCACAGAGGTAAGCCAGTGGAATTCAGAAGGCTCCGAATATGTTCA

CAGCCACAACCCCAGAACAACTCATAAACTCATTTTTGGCTAATGCATTGATT

TTCTTTGGCTTTCAAAACAAAGAATTTAAAGGNGGAAACATTCATTCACAGG

CTCCTGCATAACCACTAAGACTAAAACAGCTCCGCTTGGAACTGGGAACTGC

CCCTAAATGCTTTACATGAATATATTTCTTTTNAATCCACTTTACAATCCNTAT GAGANCN (GenBank Accession Number AQ418635)

265a14-T7

(GenBank Accession Number AQ490107)

265a14-SP6

CCTNTTAGAGGGTAGGAGGGCATCTNATCTCTTTTCAACGTAGAGTTTATGAA AAGAGCACAACAGCTTTCTGGTGGGCCACAGAGGAGGCTAGATGGCTGGNG CTTTCCGTGGAGCCATGGTCTTNNCCCTATNNNTNTGGTTTACTATNNNNNNN NTCACCACTAGAATTGCGAGGCAGACGGTGAATATNANACCTTTGCACCCAA AGATGAAGGTGCTTGCNGGTNGATGAGTTGGTT (GenBank Accession Number AQ490105)

146o10-T7

GATCATCTGTGATTAACAGAAGAGAACACCTTAAAATAATGGGGNAGAGACA AATTAACTAGGTACTAAGTCAGATTTCCAAGTGAAAAAGACATGAGCAAAAC TTCTGGTCTTAATTCCTCAAATAAAATGAAAATAAATAAAATGGTAACAGAG TGACTTTCTCTAAATTTTTTAGGGAGTTTGTTACAAACCTTTGGGCTTTACAGG CATGATTTCATGGATTCAACACGAATTACACTGATATTTAGCCTTCATTTCTC ACTTCTAAAGGGTGGGACTTACCTGCTGTGCCAGAGTAGGCTTCCAATGCTA

GTTGGTTGCTAATGTCCAT(GenBank Accession Number AQ373576)

146o10-SP6

TNTAGCTCTCTGCAGCCTAACCTCTGGTGCTCAAGTGATCCTCCCACCTTGGC

CTCCAGAGTATCTAGGACTACAGGTGTGAGATGGCTACATCTCTCTTTTTTTT

TCTTTTTGAGACAGAGTCTCACCCTGTTGCCCAGGCTAGAGTGCAATGGCATG

AGCTCGGCTCACTGCAACCTCCACCGCCTGGGTTCAAGTGATTCTCCCACCTC

AGCCTCCTGAGTAGCTGGAATTACAGGCATGCGGCATCATGCCTGGCTAATTT

TTGTATTTTTGTAGAGACGGGGTTTCAACCATGTTGGTCATGAACTCCTGACC

TATGTGATCTGCCTGCCTCACCTCTCAAAGTGCTGGGAGTCAAGCGTGAGCCA

CTGCACCCAGCCTGGCTACATTNCTTAACCTCTCTGTACCTTAAGGTNCCAAT

GTGTAACCTGTGGGCTGATAANGGANCCTTGGTGATAGAAAGTCCTGGAGGA

TGATGACCCATNTNNGTAANCCCTGNAANNGGCCNGGTTTTTAAGGCTTAAT

NCTTNGGGTTTTGGGNTTNTTTNTTTTNNAAAAGGTTTTTTTTCCNNNTTANTN

AGGTTTTGTNCTTAACNTCNTNAGNCAATAACTTNNTAACNCCNTCTTNANCC

NGNTTTTNCCCCCTNNTTGTATTTTCTTCCTTTTNNNCNNCCCCCCCCCNTA

(GenBank Accession Number AQ373574)

122k3-T7

CANCAAACATGGGTGTCACATGAGGTTCTTGCATGGAATGTGGCTGACGTTT

CCTTTGGACCATGGGACCACCACCCTCCTGGCCCATGCTTCCTTGCTGCCATC

TTTTCAGGTATGTCACCTTCCGCTGTGCAGACCACTGCTCCTGAGCGCTGGGG
TGTACTTTGGCTGTTGGGGCAACTGCAGCAGTGTCTGGGCCCACCTGCTCTTT
AGCAGCCTTAGCTTCGCTTTGGGGCCAGGCACCTCTAAGTGCAAGGCTGTGCT
AGCTACACTGTGCTTGCTGGGGAA (GenBank Accession Number AQ342807)

122k3-SP6

TGGATTTTAGATGTGTTTGTCAAATACTGATTCTTTTCACTTCTCCATGGAGTT
TTTTATTATATTAATGTTTATATTTTTCATTTCAAAACTTTATTTGTATCTTCTT
CAAATCCATATGCTCTTTCTTTCTATATGCCTTTTGTGGAGTACTACTTAGGGA
AAAGTGGGAGCAGGGAAAAGCAAAAAGAGAAAGCAGTAAGCTACAAGTCTG
TCTTTCTTCATGGTCCAAGACACATAATACTCCTGTGCAAATAACTCACAATC
TTCCCGTTCCCAACTATCACCAGACACCTGCAAGTCAGCTCACTTCAATACGG
CGTTATCAGTACTGCACAATAGTCCCCTTTTGANACAGCATANACACTATCTA
TAAAATCTCCAGCAAGCCTTTGNTTCTTTGCAGCCAGNTTCTTTTNTGGANAN
CCNGCCCATTGCTTCTTTANAAAAAAAGTTNTCNAATTTTCNTNNAAAAAANN
NNNNCTTTTTNAANNNNCCTTNGCGANATTNTNTCCCNCGTNTCNTNCNNAN
AGNNCTNAATNANCCTNAATNNNTTCTTGNACTNCCANNNNNACCNTTTCNC
NTNTCTAATTATCATTNCTTTNNNCCTTNNGTNANNCTATNNTTCCTCCCNNN
NNCATACNTNACNNNNTTTNCTTATNTNCTTTANTTNTCNTCCTACTNGTCAC
TCCTCNTTTCTNNCTCCTNTCCCCCNNCCAAAC (GenBank Accession Number AQ342805)

108c2-T7

(GenBank Accession Number AQ318321)

108c2-SP6

CAATGATCCTAGCCTTGTCACAGTTAGATTGTCCAAGTAAGGGTACCTGGCCC

AAGCTTTGCCAATTAGATACCCTNCAAGAGGAATTTGGATTAGGTTCATCCTT

GCTATGTGATATTTTGAACAACAACAACAACAACAAAAAGTAAACTCAGGCG

ATGTTGGCAGCATGCTAACTGGAAAACCAGGTGAAACTTGTCTCAGCAGATG

ACAGAGATAAGAGAGTGAACTAGTTCTAAAGAGAAGAGCAGCTGCAAGAGA

TGAATCTAAATTCTTAGGGGCATTCAAGTCCCTGCTTCCATTCATTCTGGCAA

GCTGGCTGTATCTCTGCCCTTGACTTCCATTACATGCCCCATGTATCTTTATGT

GGTGAGCTATGTATTGAGGTACCCATATTCCTGGTTATTGGANGCCAATAATC

GGAAGNAGANAACTTTTTTGGTTCNTCCATTCTCCTTANTGGCCTCCTTTACC

ATTTACCACCCCAAAANTTTTANCCAACCCAGGNTCCTTGNATCCNNTTNGNN

NNTTCATTGCTAANACCTTNGAAANCTTCNTTTTAACCTTGTCCCNNNCNTNA

NNCNTTNTCANGGNTCCCNNATNTGGAAAANNNNC (GenBank Accession Number

AQ318317)

65h9-T7

(GenBank Accession Number AQ194113)

65h9-SP6

CAAGGATCCTTGAAATCAAAATAGGATTCTACTTTAAAAACGGAAATATCAA

AAATAAAAAAAAGATTATCTTGGAAATTAAAAACACCATAACTGAAATGAAA

CAGTCAATAGAATCCTGGAAGAGAAAGTTGAAAACATCTTCCAAAAAGTAGA

GCAGAAAGATTAAAATATGGGAGAAAAGATCAGGGGTCCTATCCAGGAAGA

ACTACTGTATATCAGAATAGAAAGTGAGTAATGGAGTAAAACAGTGAGTAAT

GGAGGAGAGGAAATCAGTAACTAAACACATGAAAGAAAATCCCCAGAACTG

AAGAATATGAGTTGCCAGAGAAAGACTCATCAACTGTCTAAATGAAGCACTN

ACACCAAGCCACATCTTAGTGAAATTTTCAGAATNCTGAAGGCAACCNGAAA

ATCTCNGGCTTTTGGGAGAAACCGGTCAGTCCNANGGCCGATTCCNAANCCT

TTNNATTTANATGGCCCTTACCTTANCANNGGNAAANTNCCCCAATTCCNAG

GGAATTTTTTCCNAANAAACTCCCNNCCNTGNANTTTTTNCTNNGGGAAAAA

ACTTTCTTNTANNNNNCAAATTGCCCNCCCTTTTCCC (GenBank Accession Number AQ237184)

16f22-T7

(GenBank Accession Number B82490)

16f22-SP6

TCAGACTTCAGGTAAAAGCATCTAGATGACAAATATTTTTTAAAAATACAAA

GAGGTAAAAGATGAGTGTTTTGAATGAAGGCTGTTGCAATATGGTTTGAGAT

AAATGGATATGAGGTGTCTATTTTGCAAGTCTAGACACTTAGATGTGAAAAA

TGAGGCAAAACAAAGCATAAAAGATGACTTCTATGTTTTTAAGATGTGGGAA

TATTGACTCAAATGAGAAAGACTGGGAGCAGGACAGATTTGTGAAAGGAGC

AAATGTATTTGTGAAAGGAGCAAAAGCATCAAATCTCTGAGAACACAGGGAC

CACCGGAAAAATCCAGGTAGAAGTGAAAAACTAGAGACATAGGGACGTGGA

AGCCACTTCTGCCTGAGCACACCTGCCAGTCTAAGCAAACGTGAACTCTGATT

TTTGAACCTCATACAACACGAAAGAAATAGGTGTCAAACCCAGGACCTACTT

GAAGTGAGAATTAAGAGAAGACCTTCTCTGGCTTGTCAGACTTAAGTGATCC

TCCTTTATGGTAACATGACCTGAAATAACATGCCTACCCCGATTCCCAGGNGA

TTAAGGNAANAAAAAAATCNTTGNTAGTTTTANCCANTANCCNTGGGNTTTG

CTNNTCCNTTNCAGGCTCTAANNCANNGAATNTCCGTTCANTCTTTGAAAAA

ATTTNNNCNNNNCCTNNCCNCCCNCCC (GenBank Accession Number B85351)

Appendix 5: PAC end sequence BLASTN analysis

203o6 - T7 Nothing significant

122k3 - SP6

gb|AF001950|HSGNRHR1 Homo sapiens gonadotropin releasing ho... 80 8e-13

gb|AF001950|HSGNRHR1 Homo sapiens gonadotropin releasing hormone receptor (GNRHR) gene,

exon 1

Length = 5687

Score = 79.8 bits (40), Expect = 8e-13

Identities = 56/61 (91%), Gaps = 1/61 (1%)

Strand = Plus / Plus

146o10-SP6 Nothing significant

16f22 - SP6 Nothing significant

203o6 - SP6

emb|AL034343.17|HS108C2 Human DNA sequence from clone 108C2 on chromosome 6p12.1-21.1, complete

sequence [Homo sapiens]

Length = 130904

Score = 724 bits (365), Expect = 0.0

Identities = 437/451 (96%), Gaps = 9/451 (1%)

Strand = Plus / Minus

123h3 - T7 Nothing significant

132h3 - SP6 Nothing significant

320c12 - SP6 Nothing significant

203o6 - T7 Nothing significant

265a14 - SP6 Nothing significant

146o10 - T7 Nothing significant

122k3 - T7 Nothing significant

108c2 - SP6

emb|AL034343.17|HS108C2 Human DNA sequence from clone 108C2 on chromosome 6p12.1-21.1,

complete sequence [Homo sapiens]

Length = 130904

Score = 486 bits (245), Expect = e-135

Identities = 276/284 (97%), Gaps = 2/284 (0%)

Strand = Plus / Plus

Appendix 6: Novel STSs within the ARPKD region produced from PAC end sequences

Marker	Sequence (5' - 3')	Product Size (bp)	Annealing Temp. (°C)
203o6-SP6	(F) GCCTCCCAGAGCTGATTTCT (R) TCCACTGCTTACCTCTGTGG	178	55
203o6-T7	(F) TTTAGATCCCCTGCCAAATG (R) GCTTTAGTGAGCCCCTACCA	164	58
320c12-SP6	(F) ATGATTGGTGAGGTGCACTG (R) CCTTTTCTTGGCCCTTTTCT	207	58
320c12-T7	(F) TCCCTGACTTGCCTTCTGTT (R) GATATTGCCGGATCATTTGG	188	55
146o10-T7	(F) CCAAGTGAAAAAGACATGAGCA (R) GCAGGTAAGTCCCACCCTTT	211	52
108c2-SP6*	(F) CCCAAGCTTTGCCAATTAGA (R) GAATGGAAGCAGGGACTTGA	254	55

* The primers for the 108c2 STS were designed to incorporate a small microsatellite trinucleotide repeat which was later determined to be nonvariable in the U1 family.

Appendix 7: Subclone sequences

The ligation site between the BamHI digested pUC18 and the Sau3A digested PAC DNA produced a GATC sequence which is in bold. Any repeat sequences which were found in the subclones are in bold and underlined. Primers created around repeat sequences are underlined and a description is given in appendix 6. An asterisk (*) either before or after a GATC sequence indicates the beginning or end of the insert.

203o6 2-36 (Forward)

CTAGAG***GATC**ACGATTTGAGCATCAAGATGTTCTTGCTGGCAAAATGTTGAT

TTCTTCCCTTAAATGGTACAGATGTCATGGCAATATCAGAGAGCAAGGCATA

ATGACTTAGTGTGTGGACGCTGGAGCCAGATGCCGGGTTAAAATATCAGTTC

TGGACCTTACTAGCTGTACTAGTTGCTAATGATTTGGGTCTTGTTACCTAGCC

CCTCTGTGTCTCTTTTTCCTCATATGTAAAATAGAGATGATAATAGCGCCTAC

TTCATAAAGTCATTGAGAGAATAGAATAATAATCTATAAGGTACTTGGAATA

GTTACTGACATGGTAGGTGTATTGTTTAGTGTATTATTAAGAAAGATAAAATC

CATCTATTATTCTACCACAATAATA**GATC**ACATTAATGTTTTTCTTTCTCTTCT

GCTCATTTGCCTCACAGATATATATTTTATAAGGCTGTAACTGTAAGATACAT

GCAATTTGTATTTGGTAATTTTTAACTTCACATTATACTCTATGTATTTCCAAT

GTTTCCATAAAGTATTCACAATGATTTTTAGTAGCTGTGTAATTCTTTCAAGTT

GATGAACCAGAATTTATTTAAGTTATTCTTTTATTTTTGGACCTTTAGCTCCAA

AAATAAAAGAATAATTATTAACTTTAAAAAGTTAAATAATGCTTGTAACTGTT

TTAATTACTTTTTAAAATTATAAACTATGTGATACAAACAGAAGAGTTTATGT

CTTTGCATATGGATAGTTGCCTTGAATATCAAAATTCTGNACCGTGTGGGGGN

GGTTTCATATCTAAG

203o6 2-36 (Reverse)

CCCGGG***GATC**CTCGGCAAACGCCCGTGGTATTTGCGTCTGCGCGGCAAGCT

GTACTTGCTGCTGCAACGTAATGCGCTGCGTGCACCA**GATC**CCACTCCAACG

TTCAGCCTGCACGCCCCATCCAAGCAGTGCCTCCCAGAGCTGATTTCTGTCAA

AGCCATGANAATGTGGTTGTGGGAAATGGTTACAGGGCTTCTGGTGTTGTCA

CCACTATAAACAAAGTCTGGTCCTCACCTCAGGCCATACAAAGGGCTGCGGT

GTGTCACCTATAGATAGCACGTAACAAAACCACAGAGGTAAGCCAGTGGAAT

CAGAAGGCTCCGAATATGTTCACAGCCACAACCCCAGAACAACTCATAAACT

CATTTTTGTCTAATGCATTGATTTCTTTGGCTTTCAAAACAAGAATTTAAAGGT

GGAAACATTCATTCACAGGCTCTGCATAACCACTAAGACTAAAACAGCTCGT

TGGAACTGGGAACTGCCCTAAATGCTTTACATGTATTATTTCTTTTAATCCAC

TTTACAATCCTATGAGACAGACATTATCACCCCGCTTTATAGTTGAGGAAACT

AAGACTCAGAGAAAATTAACTAATTTGCCCCAAGGTAGGGCCATTGATNTTA

AAA**CACACACACA**CGTTCCAGAATTTTTGATTNTTCAANGNAACTATTCCAT

ATGCAAAAGACATAAACCCTTTTTGTTTGNATTAACAATAGNTTATTAATTTT

AAAAAAGTAATTTAAAAAAGTTCCAGCCTTTTTTNAACTTTTTAAAGGNANA

ATNTTCTTTTTTTTTGGAGCTAANGNNCAAAAATAAAAAATTACTAATAAATT

TTGGTTATAACTTNAAAA

265a14 1-3 (Forward)

CTAGAG***GATC**ATATGCACTGACTNAGAAAAGGAATGCACATGGGATGGCAA

AATAGAGAAGGCTTNTTGTGAGTGTGTGGGGGAGGCAGGATAGGGTTATCTG

GGTCTTTAAAAAAACACTTGACTTAAGCAGAGAGGGCTATGAGAGGGCTTAA

TAAAAGTACAGAGGTAGAAATACAGAAGAGGTATAAAAATAACAAATAGTC

TTATGTGATTACAACTTGACATATGTAATTTAGCGGAAGCATTTTCTACTTCT

GAAACAAATTATGATTAACACTTCCTGTAATAGGTTTATAATTATGTAAACCT

ATTTAAATCAATCATTTGCACAACAAAATGAATGAATTAATTTGTTGCAGCTG

TGTTTACTGAGGCACACATGTAAAGCAATTTTAAATGTTGTCAAGTAACAATG

ATTTCGTTTCATGACTCACACTAAGTGAGAAATTCTGATTTTTTTAAGACAAA

GTTTAAAAAATCAAAAACTTAAAACTCCTGGAACTAAAAGCATAGAAGAAAA

GCTTCATGACATAGGACATGACAATGTGTTCTCACATATGACACCAAAGCAC

AGGGAAAGAAAGCAAAAGTAGAAAAGGGACTATATCAAACTCAAAAACTTT

TGTGCATTGAAGGACACACCTAACAGCCTGAAAAGGCAACCTATGGAGTAGG

AGAAAATATTTTGCAAATCATATATATCTATACATATATATGGATATGATAGT

AAATTGTTATACAGAAAATNTAAAAAGACTTNTNTTACTNAATAACAAAAAT

CAAATACTTATTTTTTTAAATGGCTAA

265a14 1-3 (Reverse)

CCCGGG***GATC**CCAGAGGAATCACCAGACTGTATTTCACAATGATTGGACTA

ATTTGCACTCCCACCAACAGTGTAAAAGCATTCCTATTTCTCCACATCCTCTC

TAGCATCTGTNGTTTGCTGACTTTTTAAT**GATC**AAGGAATAGCAAAGCTTTGA

GAGTGGCACCTGCTTTGTGGAAATGATACCTACAATTTTAATTTCCTTTATCT

CACTCTACTAGCATTTCCTAACCAATTTATAGAAACTGACATTAAATAAAGTT

AAAGTGAAGAGCTGGCCTATCAGTCAT**TGTGTGTGTGTGTGTGTATGTGTG**

TGTGTGTTGCACATACCANATGTATATACCAAGAAGTGTTTGTATTCTTTAA

ATCTTGGATTTTATACTGAAGGTTATAATGACCCGGATTATCTGAATGTTTAC

ATTTAAAAATGTGGCCAGGAGACTTTAAGCAATGAGAGTCCTAATGCAAATG

TGTGATGAAAACAAAATCATGACAATGTTGTNTTAGAGATTCATCAAAGCAC

AAACGAACATTGGCTCTTGATAGATGATGTACATTGGGTCAATGCTAGCCTTT

TGATTTTCACCTCACATATTTATCCATCGNAAACATGAANGGGGTAAGCTTTC

CTGCTTGTGTAGAAATAGNCAGGAACACCTGGGATTAACTTCNGATGAAGAC

NCCTAAATGNGNAATTNGGAAGGGTAAAATTTTATTTTGGCTGGGTAATNTN

AAACCCAGCNCCNTCTTCTAAAGGGCCNAGGNCCTTTTTTGGTNAAAATGGG

NAANTTGGGTTTAAACTTTTTAAAGCCCATCCCTTTCCTAAATTTGGGG

203o6 4-22

This clone gave a clean sequence for the pUC18 vector at both ends but the sequence data clearly indicated two insert sequences present. This indicated that the bacteria were

transformed with two separate ligation products. A clean sequence was not attained.

16f22 2-22 (Forward)

CTAGAG*GATCACGCCAGATGCAGTATGTGTGTGAGAGAGAGAGAGAGACA

GAGTGTGTGAGAGTGTGTGCAAATGTGTGTGAGCATGTGTAAACGCATGCAT

GTGTGCGTGCATGTGAGTGTATGCATGTACGAGTGTGAGAGTACGTGCAGAT

GCATGTGATTGTGTGAGATTGTGTGTGCATATGTGAGCATGTGTGAGAACATG

TATAAGTGCATGCGTGTGTGCATGTGTGTGAGTGCATGTATACGTGTTTGAGA

GTATGTGTGCACCTGCATGTGATTGTGTGTGTGAGAGAGTGTGAGAAACAAA

GTGTGAATGTGTGCATATGTGAGTGTGAGAGCATGTGTAAGTGCATGCGTGT

GTGTGGATGTGTGTTTGAGAGTATATGTGCACATGCATGTGATTGAGATAGA

GTGAGACAGTGTGCATATGTGAGTGATAACGTGTAAGTGCATGAGTGTGAGT

GTGCGTGGATGTGAGAGTATGTGTGCACATGCATGTGATTGTGTGTCTGAGAC

TGTGAGCATGTGTGTGTGGGTGCATGTGTATGTGGGTGCATGTTGTGTGTGTG

AATGTGTGTTAGAGCATGTGTGGGTGCATATGTGAGACTGCATGTGTGTGTGA

GTGCATGCATGTTTGGGAGACAGCATGTGTGGGTTCATGCATTNGAGCACGT

AGTGTGTCAGTGTGTGCATGTATCCCGNGTGNGNGNGNGNNN

16f22 2-22 (Reverse)

CCCGGG*GATCATCTCTATACAAGTCATTTGCTAAGTTTTATGAAAGCCAGAG

AGAACAAATCATTCTAATAGATTGTGGAAAAGGTTTTATAGAAAAAGTGGGG

GGAAAACAGAGGCGTGACAATGCGGGTGTTTCAGGAAATGGTGTGACTACAG

TCTAGTGGGTGGGGAAGTGACCGAAAATGAGACTGAGTCTGGACTGAAAAGT

GACCTGAGTGACACACATAAAGGAACTGAAACGTAGCAATCAGAAACCGCG

GCGTCATGCTAACTTCTGTAACTCCAGCACCCAGCACAATTCCTCAGGCAAA

GCAGGCTCTCAGTAACGCCTGTGACGTTAACACATTCCTGAGAGGAACATGG

AGCCGCCGAGGAGCTGATGGAGAAAGCGGTGCACACAACTTCGCGTTTTGGA

AAGATAATCTTGGTCACTGTACAGAACAAACTGGAGTGGCAAGGACATGGAA

GACCTCCATCAGAATTGGAAAAATTAATTTCATCTATTATGTCCTATTATGTC

CTAATGACAAGATTTTCCCCTGCAGGCACCAGCAAAAATCAAACTACAAATG

CCGGGCTAAACCAAGTCCAGAGAATTCCTGAGGTCCTGAGGACAGGGTCCCC

CTAAGTACTCCTCCTGCATTCATATCTTCTGTCTCAAGCCTGTGTTNCTTTTAA

ACACACACACACAACACNGGNTTCCTTTCCCCACACTGGCCACACTTCACCG

TGCCTCAAAATGGCATGGAACCCCCCCCATGCTTTTTTCCCAAACATGCCATG

CACCTTCACACCACACCATTCCAGTT

16f22 1-40 (Forward)

CTAGAG***GATC**TGTTTCTATTTAATACTTAGAGAATTGTAGGTTCTAGGTACT

CTGAAACAGCCACTAATTTCATTCCCTTGGCTTACCTATTCCCTTGGATATAG

GTAAGACATTTCATATATTTGTATTTT**GTGTGTGTGTGTGTGCATGTGTGTG**

TGTGTCTGTCTGTCTGTCTGTCTTGTGGCGCTTTAATATTATTCTGGCAT

TTTAGAATGGATAAATTTCTATAATTTACTCTCCA**GATC***CCCGGGTACC

65h9 4-34 (Forward)

CTAGAG***GATC**GCCGAGCGCCCCGGCAAGGTAAGAACCTTGAAACAGCATCC

ACGCAAGAACAAAACGGCCATCAACATCGAATACATGAAAGCCAGCATCCG

GGCCAGGGTGGAGCACCCATTTCGCATCATCAAGCGACAGTTCGGCTTCGTG

AAAGCCAGATACAAGGGGTTGCTGAAAAACGATAACCAACTGGCGATGTTAT

TCACGCTGGCCAACCTGTTTCGGGCGGACCAAATGATACGTCAGTGGGAGA**G**

ATCTTGGCCCCTGCCCTCCTCTTAAATCGCCATCCCCTACCCCCAACACCTTA

AAAAAATGATAATAAAGTACATGAAGACAGTTTGAGAGCTTTAGGTTTCCCA

CTGAGGTTTGGAGCTTCTCTCATGGGAACCGGTCCCCAGGCTGCATTCACGAA

TTGCTCTCATCTCCATTCTCCCACAGGTTTGTTCTCCACCCGTACCCCCCACCT

CGCCCGGCTATCTTACTCCCTGCGATTTCTATAGGAAAAAGTGTGTTCTTGGC

AGGGCGTGGTGGNTCATGCCTGTNATTCCAACACTTTGNGAGGNCNNGGCNG

NCG**GATC**GAC

65h9 4-34 (Reverse)

CCCGGG***GATC**CCTCTAGGCATTTGGGAGTGAGGGGTTCTGCTCCTCCTCAGC

TCTAACTCCGTGCTAGTGGGGGGAGAGACATAATAAACAAGTAAACAAATAC

ATATGATTTTTTACAAAAATTATAATAAGTTCTATGAAAGAAATAAAGCAGG

GAACTGAGGCAGAGGCAGCATTCTAGGATAGGTAGGATAGGT**GATC**AAAAG

TGGGCAATATTGCCTCTAGAGAGATGTGGCAGCAAGGACATTTTAAGTAAAG

GCAGACATGGCCAAGCCCTTAGGAGATACTAATGTGACTAGAATGAAGGGTC

CACAGGAAAGGTGAGTACAGGGTACGACTGAAAGCAGCTGGGTACATTGGG

CCTGACTCTAGAGGTCACAGCCAACAGAATAAGTACACAGTGCCTGATTTAG

TAGGAAAAAGGGAGCCATTGATGGCTTTTGAGCAAAGGAGAGACTATCAAA

GTTGTGCTTTAGGAACATTAATCTGGCCAGAGCATGTGTAGAATTAATTATTT

CAATCAAATAAAAGGAAATGAAGCTGGGGCCAGGCTCAGTGGCTCATGCCTG

TAATTNCAGCACTCTGGGAGGCTTGAGGTGGNTGGNTCCTNAAANNNAATAC

TCCTTTTTNTTATTANGCCTTTC

65h9 1-9 (Forward)

CTAGAG***GATC**AAAAGCACAGAAAGGTGTCCACTTCTTACCCTGTTGGAAGG

CACAAAGGATGGTTTAGAGGGAGGTATGTTTATGTCTGCCTCACTTATTGAAA

AATACATCTGTACTTATGCCCTCCACTTTCTCATCAACTATTTCTATGCATTCT

GGCTTCTTCTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTCACTCTGTCGCCC

AGCTGGAGTGCAGTGGCAT**GATC**CTCCCACCTCAGCCACTTGAGTAGCTTGG

ACTACAGATGCGTACCACCATGCCTGGCAAATTTTTTTTTCTTTTTAAGTTTTT

ATAGGGACAGGGTTTTGCATGTTGCCCAGGCTGGTCTCGAACTCCTGGGCTCA

AGCAGTCCACCTGCCTCAGACTCCCAAAGTGCTGGGATTACAGGTGTGAGTG

GATGTGCCTGGCCAGTTTCTAAGATGTTATGATATATA**GATC**ACAAGGTCAG

GAGCTTGAGACCACCCTGACCAACATGGTGAAACCCCATCTCTATTAAAAAT

ACAAAAGAAAAATTAGCCCGGTGTGGTGGCGCATGCCTGTAATCCCAGCTAC

TCAGGAGGCTGAGGCAGGAGAATAGCTTGAACCCGGGAGGCAGAGGTTGCA

GTGAGCCAA**GATC***CCCGGG

146o10 5-37 (Forward)

CTAGAG***GATC**CCTCCATTTGGGGTCCCTGACTTCCCGCAACTCTGCTAAAGG

TAGTGTATTAACATCCACACAAAGCCCAGAAAAAAAATGAACAAAGACTGAA

GCAAAATGTGCCTTTAGGCCAATCTCAAAAGAAACTTAAGAACATTGGGGGG

TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGGTGGGGTTTTGTTGT

TGTTGTTTGTTTTTTTGAGACAAGGTTTCACTCTGTTGCCTAGGTTGGAGTGCA

GTGGCAT**GATC***CCCGGG

Appendix 8: STS primers designed around novel satellite repeats found in PAC subclones

Marker	Sequence (5' - 3')	Product Size (bp)	Annealing Temp. (°C)
265a14 1-3	(F) TGAAGAGCTGGCCTATCAGTC (R) AAGTCTCCTGGCCACATTTTT	177	55
16f22 1-40	(F) TGGCTTACCTATTCCCTTGG (R) CCGGGGATCTGGAGAGTA	174	55
146o10 5-37*	(F) TGTGCCTTTAGGCCAATCTC (R) GGATCATGCCACTGCACTC	163	??

*Unfortunately the Reverse primer of the subclone 146o10 5-37 included part of the vector which may account for the lack of a product when used to analyze the U1 DNA.

Appendix 9: PAC subclone sequence BLASTN analysis

203o6 2-36(F)

emb|AL034343.17|HS108C2 Human DNA sequence from clone 108C2 on chromosome 6p12.1-21.1, complete

sequence [Homo sapiens]

Length = 130904

Score = 1029 bits (519), Expect = 0.0

Identities = 526/527 (99%), Gaps = 1/527 (0%)

Strand = Plus / Plus

203o6 2-36(R)

emb|AL034343.17|HS108C2 Human DNA sequence from clone 108C2 on chromosome 6p12.1-21.1, complete

sequence [Homo sapiens]

Length = 130904

Score = 1025 bits (517), Expect = 0.0

Identities = 526/528 (99%), Gaps = 1/528 (0%)

Strand = Plus / Minus

emb|X68551|ECKUPRBS E.coli kup gene for integral membrane protein

Length = 2367

Score = 163 bits (82), Expect = 1e-37

Identities = 82/82 (100%)

Strand = Plus / Plus

gb|AE000452|AE000452 Escherichia coli K-12 MG1655 section 342 of 400 of the complete genome

Length = 10700

Score = 155 bits (78), Expect = 2e-35

Identities = 78/78 (100%)

Strand = Plus / Plus

gb|L10328|ECOUW82 E. coli; the region from 81.5 to 84.5 minutes

Length = 136254

Score = 155 bits (78), Expect = 2e-35

Identities = 78/78 (100%)

Strand = Plus / Plus

265a14 1-3 (F)

gb|AC005886.2|AC005886 b240g16, complete sequence [Homo sapiens]

Length = 134011

Score = 56.0 bits (28), Expect = 1e-05

Identities = 40/44 (90%)

Strand = Plus / Minus

emb|Z97180.1|HS203P18 Human DNA sequence from PAC 203P18 on chromosome Xq27.1-Xq27.3.

contains CpG island, and polymorphic CA repeat

Length = 132470

Score = 54.0 bits (27), Expect = 6e-05

Identities = 42/46 (91%), Gaps = 2/46 (4%)

Strand = Plus / Minus

265a14 1-3(R) Nothing significant

16f22 2-22(F) Nothing significant

16f22 2-22(R) Nothing significant

16f22 1-40(F) Nothing significant

65h9 4-34(F)

emb|AL034343.17|HS108C2 Human DNA sequence from clone 108C2 on chromosome 6p12.1-21.1, complete

sequence [Homo sapiens]

Length = 130904

Score = 547 bits (276), Expect = e-153

Identities = 309/325 (95%)

Strand = Plus / Minus

65h9 4-34(R)

emb|AL034343.17|HS108C2 Human DNA sequence from clone 108C2 on chromosome 6p12.1-21.1, complete

sequence [Homo sapiens]

Length = 130904

Score = 674 bits (340), Expect = 0.0

Identities = 340/340 (100%)

Strand = Plus / Minus

65h9 1-9(F)

emb|AL034343.17|HS108C2 Human DNA sequence from clone 108C2 on chromosome 6p12.1-21.1, complete

sequence [Homo sapiens]

Length = 130904

Score = 361 bits (182), Expect = 2e-97

Identities = 182/182 (100%)

Strand = Plus / Minus

146o10 5-37 Nothing significant

Appendix 10: Sputnik output for the 108c2 PAC clone full sequence.

The output in bold indicate several separate repeats which all occur as a single larger complex repeat region. It was this region for which the 108c2 FULL1 primers were created.

>gi|5931807|emb|AL034343.17|HS108C2 Human DNA sequence from clone 108C2

trinucleotide 119 : 138 -- length 20 score 17

AACAACAACAACAACAACAA

tetranucleotide 3973 : 3985 -- length 13 score 9

CCCTCCCTCCCTC

tetranucleotide 3990 : 4009 -- length 20 score 16

TCCTTCCTTCCTTCCTTCCT

trinucleotide 5226 : 5237 -- length 12 score 9

AGGAGGAGGAGG

Dinucleotide 6158 : 6176 -- length 19 score 17

ACACACACACACACACACA

tetranucleotide 7102 : 7115 -- length 14 score 10

ATAAATAAATAAAT

pentanucleotide 7272 : 7289 -- length 18 score 13

AAATAAAATAAAATAAAA

Dinucleotide 8723 : 8755 -- length 33 score 31

ATATATATATATATATATATATATATATATATA

Dinucleotide 11881 : 11894 -- length 14 score 12

GAGAGAGAGAGAGA

pentanucleotide 13675 : 13688 -- length 14 score 9

AAAGCAAAGCAAAG

tetranucleotide 14662 : 14736 -- length 75 score 59

TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTC

TTTCTTTTTCTTTCTTTCTTTCT

tetranucleotide 14737 : 14765 -- length 29 score 25

CCTTCCTTCCTTCCTTCCTTCCTTCCTTC

Dinucleotide 14768 : 14787 -- length 20 score 18

TCTCTCTCTCTCTCTCTCTC

tetranucleotide 14788 : 14835 -- length 48 score 44

CTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC

Dinucleotide 14836 : 14912 -- length 77 score 33

TCTCTCTCTCTCTCTCTCTCTCTCTCTTTCTCTCTTTCTCTCTTTCTCTC

TTTCTCTCTTTCTCTCTTTCTCTCT

tetranucleotide 14913 : 14929 -- length 17 score 13

TTCTTTCTTTCTTTCTT

pentanucleotide 22294 : 22312 -- length 19 score 14

TTTTGTTTTGTTTTGTTTT

tetranucleotide 26718 : 26740 -- length 23 score 12

TTTGTTTGTTTGTTCGTTTGTTT

Dinucleotide 29477 : 29487 -- length 11 score 9

AGAGAGAGAGA

tetranucleotide 30893 : 30915 -- length 23 score 19

TTTATTTATTTATTTATTTATTT

tetranucleotide 33163 : 33177 -- length 15 score 11

AATGAATGAATGAAT

pentanucleotide 33968 : 33981 -- length 14 score 9

TTTTATTTTATTTT

Dinucleotide 37953 : 37966 -- length 14 score 12

GTGTGTGTGTGTGT

tetranucleotide 40200 : 40212 -- length 13 score 9

TTTCTTTCTTTCT

trinucleotide 40434 : 40453 -- length 20 score 17

TTCTTCTTCTTCTTCTTCTT

pentanucleotide 40725 : 40739 -- length 15 score 10

AGGGGAGGGGAGGGG

trinucleotide 40755 : 40766 -- length 12 score 9

GGAGGAGGAGGA

pentanucleotide 43320 : 43344 -- length 25 score 13

AAAACAAAACAAAAACAAAACAAAA

pentanucleotide 52428 : 52445 -- length 18 score 13

TTTTCTTTTCTTTTCTTT

tetranucleotide 53952 : 53979 -- length 28 score 17

ATTTATTTATTTATTTACTTATTTATTT

trinucleotide 54272 : 54283 -- length 12 score 9

TTATTATTATTA

trinucleotide 56825 : 56838 -- length 14 score 11

TTGTTGTTGTTGTT

tetranucleotide 59897 : 59911 -- length 15 score 11

TTTATTTATTTATTT

trinucleotide 63337 : 63349 -- length 13 score 10

GAAGAAGAAGAAG

tetranucleotide 66349 : 66370 -- length 22 score 18

TTATTTATTTATTTATTTATTT

pentanucleotide 73277 : 73295 -- length 19 score 14

TAAAATAAAATAAAATAAA

pentanucleotide 76547 : 76560 -- length 14 score 9

TTTTGTTTTGTTTT

pentanucleotide 81811 : 81824 -- length 14 score 9

CTGTGCTGTGCTGT

pentanucleotide 83986 : 83999 -- length 14 score 9

TTTTGTTTTGTTTT

tetranucleotide 87559 : 87573 -- length 15 score 11

TTTCTTTCTTTCTTT

pentanucleotide 91317 : 91346 -- length 30 score 25

CAAAACAAAACAAAACAAAACAAAACAAAA

pentanucleotide 91692 : 91724 -- length 33 score 28

AAACAAAACAAAACAAAACAAAACAAAACAAAA

Dinucleotide 95163 : 95183 -- length 21 score 19

CTCTCTCTCTCTCTCTCTCTC

Dinucleotide 95185 : 95200 -- length 16 score 14

ATATATATATATATAT

Dinucleotide 96248 : 96260 -- length 13 score 11

AGAGAGAGAGAGA

pentanucleotide 100016 : 100029 -- length 14 score 9

CTGAACTGAACTGA

tetranucleotide 104791 : 104803 -- length 13 score 9

TATTTATTTATTT

tetranucleotide 107121 : 107152 -- length 32 score 28

AAACAAACAAACAAACAAACAAACAAACAAAC

tetranucleotide 108459 : 108489 -- length 31 score 13

TTCATTCATTCACTCATTCACTCATTCATTC

trinucleotide 110705 : 110723 -- length 19 score 16

TAATAATAATAATAATAAT

Dinucleotide 112705 : 112717 -- length 13 score 11

TCTCTCTCTCTCT

Dinucleotide 112720 : 112746 -- length 27 score 18

ACACACACACACACGCACACACACACA

pentanucleotide 114775 : 114788 -- length 14 score 9

TATAGTATAGTATA

pentanucleotide 120358 : 120372 -- length 15 score 10

CAATTCAATTCAATT

tetranucleotide 122465 : 122478 -- length 14 score 10

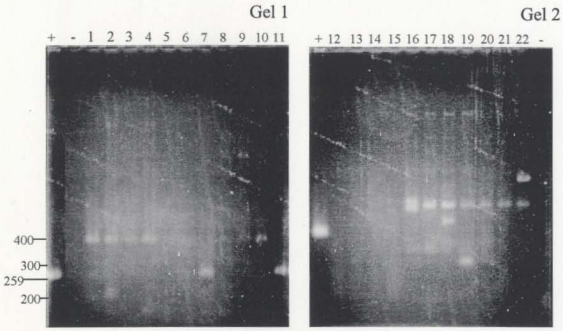
AAATAAATAAATAA

tetranucleotide 126166 : 126209 -- length 44 score 26

TCCCTCCCTCCCTCCCTTCCTCCATCCCTCCCTCCCTCCCTCCC

trinucleotide 127561 : 127574 -- length 14 score 11

TCTTCTTCTTCTTC



1. 1-10	7. 61-70	13. 121-130	19. 181-190	25. 241-250	31. 301-310
2. 11-20	8. 71-80	14. 131-140	20. 191-200	26. 251-260	32. 311-321
3. 21-30	9. 81-90	15. 141-150	21. 201-210	27. 261-270	
4. 31-40	10. 91-100	16. 151-160	22. 211-220	28. 271-280	
5. 41-50	11. 101-110	17. 161-170	23. 221-230	29. 281-290	
6. 51-60	12. 111-120	18. 171-180	24. 231-240	30. 291-300	

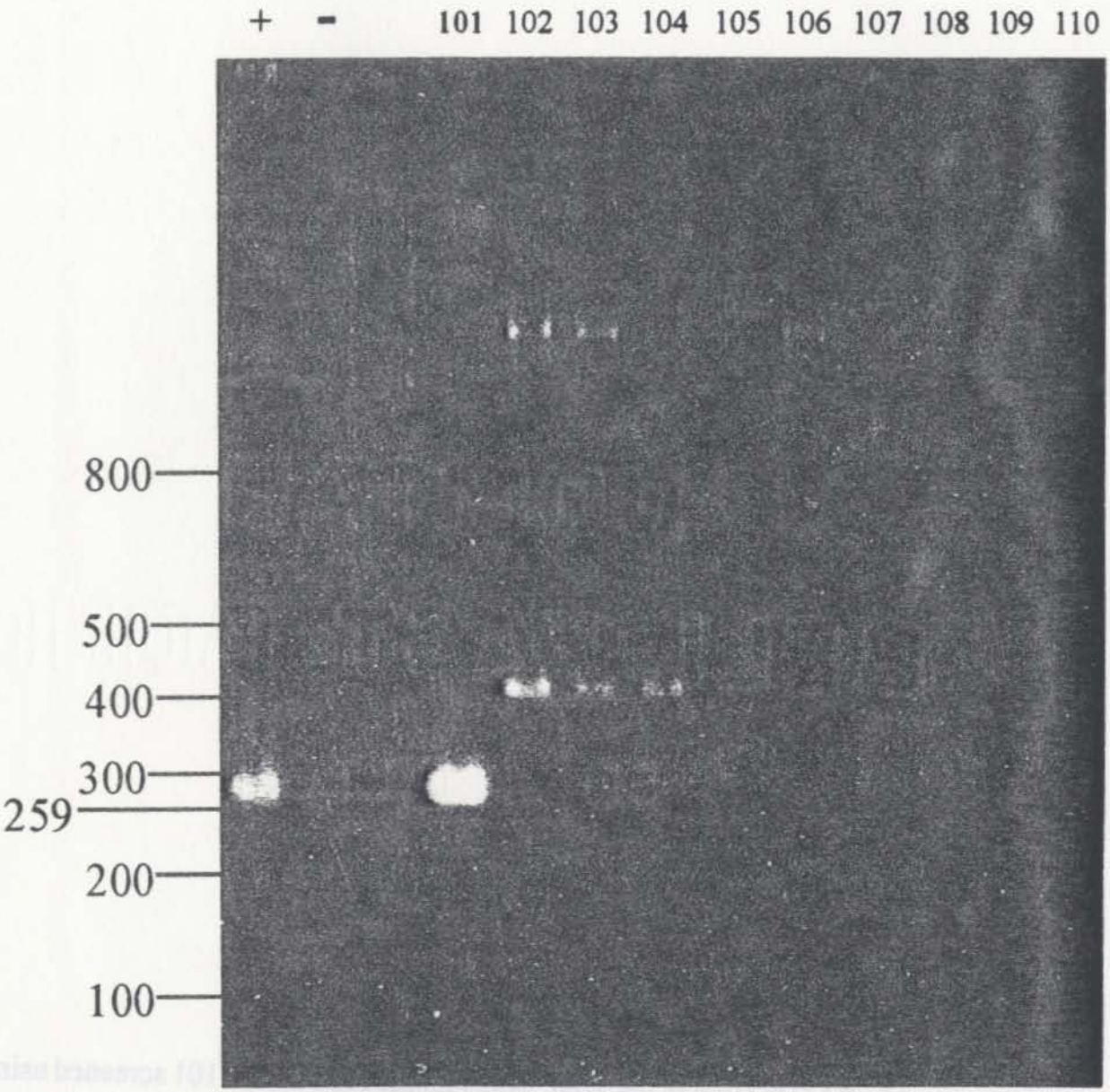
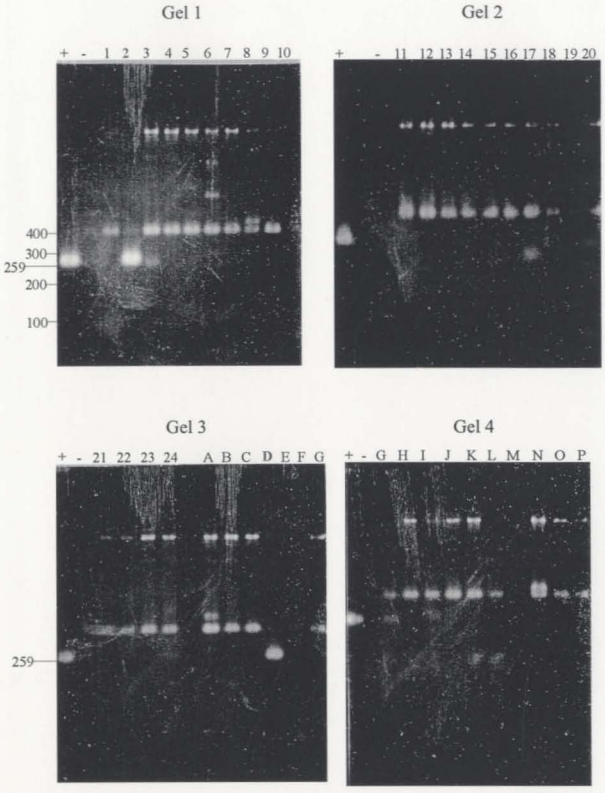


Figure 11: Plate pool screening of the PAC "Plate Pools" with the marker 855R. Note that the marker amplified a fragment in 101, indicating that the PAC of interest is in that plate. + = positive control (human genomic DNA), - = negative control (dH_2O).


Gel 1
+ - 1 2 3 4 5 6 7 8 9 10
400
300
259
200
100
Gel 2
+ - 11 12 13 14 15 16 17 18 19 20
Gel 3
+ - 21 22 23 24 A B C D E F G
259
Gel 4
+ - G H I J K L M N O P