European Journal of Human Genetics (2000) 8, 163–166

www.nature.com/ejhg

ARTICLE

Genomic structure of the gene for the human P1 protein (MCM3) and its exclusion as a candidate for autosomal recessive polycystic kidney disease

Yvonne Hofmann[1], Jutta Becker[1], Felicia Wright[2], Ellis D Avner[3], Michal Mrug[4], Lisa M Guay-Woodford[4], Stefan Somlo[5], Klaus Zerres[1,6], Gregory G Germino[2] and Luiz F Onuchic[2]

[1]Institute for Human Genetics, University of Bonn, Germany; [2]Department of Medicine, Johns Hopkins University, Baltimore, ML; [3]Department of Pediatrics, Rainbow Babies' and Children's Hospital, Case Western Reserve University, Cleveland, OH; [4]Department of Medicine and Pediatrics, University of Alabama at Birmingham, AL; [5]Yale University School of Medicine, New Haven, CT, USA; [6]Institute for Human Genetics, Technical University of Aachen, Germany

The locus *PKHD1* (polycystic kidney and hepatic disease 1) has been linked to all typical forms of the autosomal recessive polycystic kidney disease (ARPKD) and maps to chromosome 6p21.1–p12. We previously defined its genetic interval by the flanking markers D6S1714 and D6S1024. In our current work, we have fine-mapped the gene for the human P1 protein (*MCM3*), thought to be involved in the DNA replication process, to this critical region. We have also established its genomic structure. Mutation analyses using SSCP were performed in ARPKD patients' cDNA samples, leading to the exclusion of this gene as a candidate for this disorder. We also identified two intragenic polymorphisms that allowed families with critical recombination events to be evaluated. Although neither marker was informative in these individuals, they are the closest yet described for PKHD1 and may help to refine the candidate region. *European Journal of Human Genetics* (2000) **8**, 163–166.

Keywords: ARPKD; PKHD1; *MCM3* gene; P1 protein; polymorphic markers

Introduction

Autosomal recessive polycystic kidney disease (ARPKD) is one of the most important inherited renal diseases in the perinatal period and early childhood. Its basic features include fusiform dilatation of the kidney collecting ducts and biliary disgenesis associated with portal fibrosis.[1,2] ARPKD is associated with a high perinatal and immediate post-birth mortality rate and often leads to a need for dialysis or kidney transplant, as well as to episodes of esophageal varices bleeding and portal thrombosis.[3] Its locus, *PKHD1* (polycystic kidney and hepatic disease 1), was linked to chromosome 6p21.1–p12.[4,5] The interval of interest was progressively narrowed by recombination analyses to the current flanking markers D6S1714 and D6S1024.[6]

The gene for the human P1 protein (*MCM3*; GDB: HSP1H), originally mapped to 6p12, is also known as Replication Licensing Factor beta subunit gene (GDB: HUMHRLFB). Its cDNA encodes a nuclear protein of 105 KDa and is 60% homologous to the yeast *MCM3* (minichromosome maintenance deficiency) replication control gene of *S. cerevisiae*.[7] It is thought to be involved in the control of single DNA-replication per cell cycle in association with the replication-specific DNA polymerase alpha. *MCM3* is highly conserved in eukaryots and is ubiquitously expressed. Considering that a mutation in *MCM3* could potentially lead to loss of function related to the regulation mechanism, double DNA replication per cell cycle, with cell proliferation as a consequence, could occur. Given the likely developmental nature of the basic defect in ARPKD and the hyperplasia observed in the

Correspondence: Klaus Zerres MD, Institut für Humangenetik, Universitätsklinikum der RWTH Aachen, Pauwelsstraße 30, D-52074 Aachen, Germany. Tel: +492418080178; Fax: +492418888580; E-mail: kzerres@post.klinikum.rwth-aachen.de
Received 28 June 1999; revised 28 September 1999; accepted 30 September 1999

collecting duct and biliary epithelia, abnormalities in regulation of the DNA replication process could potentially lead to the disease. Altogether, if placed in the region of interest, *MCM3* should be considered a significant candidate for *PKHD1*.

In the current study, we mapped the *MCM3* gene to the disease region and evaluated its candidacy to *PKHD1* via mutation analysis in ARPKD patients. The genomic organization of MCM3 was analyzed and exon–intron boundaries were identified.

Results and discussion

We have previously cloned the PKHD1 region in a set of overlapping YACs[8] and more recently have converted this into a PAC/BAC-based contig spanning the interval.[9] Database search (http://www.ncbi.nlm.nih.gov/UniGene) revealed that the sequence tag site (STS) stSG24961 (Sanger Centre), contained within the AA232139 and several other overlapping expressed sequence tags (ESTs, http://www.ncbi.nlm.nih.gov/genemap98/loc.cgi?ID = 12338), mapped between D6S427 and D6S1623 (http://www.ncbi.nlm.nih.gov/genemap98/map.cgi?MAP = GB4&BIN = 210&MARK = stSG24961), an interval that includes the PKHD1 region.[6] Analysis by BLAST revealed that these stSG24961-related ESTs were part of the *MCM3* cDNA. The STS stSG24961 is located between bp No. 2815–2935 of the transcript (Acc. No. D38073), and was mapped into the PKHD1 interval by PCR using the information provided by the YAC and PAC/BAC contigs.[8,9] The expected size product was detected for all the YACs 802d11, 961c3, 982f3, 855e9 (CEPH YAC library),[8] for the PACs, 203O6, 489K19, 406B13, 729O24, 65H9, 363P18 (RPCI PAC genomic library) and for the BAC 679J14 (RPCI-11 BAC genomic library).[9] These data confirmed *MCM3* within the critical interval, placing it between the STSs P489K19sp6 and P363P18T7.[9]

The human *MCM3* cDNA sequence was first released to the database as GenBank locus HSP1H, accession No. X62153, consisting of a cDNA sequence of 2575 bp which included the entire proposed coding region.[7] Another cDNA was independently cloned later having been called human Replication Licensing Factor beta subunit gene (GenBank locus HUMHRLFB, accession No. D38073).[10] This cDNA, longer than the previous one (3071 bp), also contained the entire coding region. Alignment of these cDNAs revealed that they are about 99% identical. Whilst the HSP1H cDNA presents several mismatches when aligned to the genomic sequence contained in the PKHD1 interval, the HUMHRLFB cDNA sequence matches it. The only two mismatches originally detected between the HUMHRLFB coding region and the genomic sequence were proved to be sequencing errors in the HUMHRLFB cDNA sequence. In addition, our mapping results could not detect the existence of MCM3 homologs within the critical interval. Altogether, these data strongly suggest that HSP1H and HUMHRLFB are the same gene, and that the right cDNA sequence is HUMHRLFB.

The *MCM3* cDNA and markers from the BAC/PAC contig were used to search the Sanger Centre Human Chromosome-specific BLAST server (http://www.sanger.ac.uk/HGP/Chrom—blast—server.shtml). The 108C2 PAC insert sequence was pulled as unfinished sequence. This sequence was generated by shotgun strategy and spans approximately 131 kb. It is currently provided by the server as a continuous sequence. The entire MCM3 cDNA was found to be contained within the dJ108C2.01299 contig. The exon–intron junctions and the 5' to 3' orientation of the genomic sequence were established by aligning the *MCM3* transcript and the 108C2 sequence using the NCBI Blast server (http://www.ncbi.nlm.nih.gov/gorf/b12.html). Putative exon–intron boundaries were assigned by direct evaluation of the break points between the cDNA and the genomic sequences, based on the splicing consensus sequences (Table 1). Given that the available cDNA sequence includes the entire coding region and is close in size (about 3.1 kb) to its message (about 3.2 kb),[7] our analysis most likely indicates that the *MCM3* gene consists of 17 exons and 16 introns and spans at least some 20.7 kb of genomic sequence (Figure 1).

SSCP analysis was performed in 23 ARPKD-patients and seven controls. All cDNA fragments in both ARPKD patients and controls showed the same SSCP patterns, with the exception of two adjacent and partly overlapping *MCM3* fragments (Figure 2). Subsequent studies on these fragments revealed a bi-allelic polymorphism in the overlapping region at position 2329 of the coding region that has a heterozygosity frequency of 51% (P1/MCM3–A/B marker). There is a G–A transition that results in substitution of lysine (AAA = allele 1) for glutamic acid (GAA = allele 2) at position 777 of the aminoacid sequence. Although this is a non-conserved amino-acid substitution, this variant was found in both affected individuals and normal controls, excluding it as a pathogenic change.

Analysis of genomic DNA of 20 individuals led to the identification of an intronic polymorphic marker in intron 2, named P1/MCM3–C (primer sequences: F: 5'-CAGGATTGATCCACAAGCCC-3' R: 5'-GACTCAGTAAAATACCAGAA-3'). This microsatellite marker consists of a triple A deletion (allele 2) within a polyA sequence of 18 bp (allele 1) starting at position 137 of intron 2 (Figure 2). Primers were designed within intron 2, yielding a PCR product of 185 bp. The study of this intronic marker in ARPKD families and control individuals showed heterozygosity frequency of 40%.

The identification of these two new polymorphic markers and the exclusion of *MCM3* as a candidate for *PKHD1* led us to perform haplotype analyses in the key recombinant ARPKD families (figured in Muecher *et al*[6]), in an attempt to narrow down the region of interest. Both markers (P1/MCM3–A/B and P1/MCM3–C), however, were not informative in any of these published families. Although not informative in such families, they may be useful if other

Table 1 Exon–intron boundaries of the *Mcm3* gene

Exon	Exon length (bp)	Part of previous intron 3' splice acceptor (AG)	Exon	Part of next intron 5' splice donor (GT)	Intron length (bp)
1	78[a]		ATGGCG...GACGAG	**gt**gagggaggcgccggcgcgcgg	1190
2	113	gtgatatgtggtggctgttttgc**ag**	GAAGAC...TAACCG	**gt**gagaggcaaggggatatactg	432
3	209	ttttctttctcttgcttatgtac**ag**	GCTTCT...CTAAAT	**gt**aagtgggttaaaggggccgct	477
4	131	atatatgcacattttccatccct**ag**	GTTCTC...ACCAAG	**gt**aaggggataaatcagttgag	2445
5	239	tgtccctctgtgtgccttttctc**ag**	GATGAG...CTTCAG	**gt**aaagggtctgccttaaagtatat	510
6	109	ttcctctttattctgtttgtccc**ag**	GACTGT...TCCAAG	**gt**ctgggcatcctttagaaatttga	1053
7	154	ttgccattggttatttctcctgt**ag**	GATATC...TAATAG	**gt**acgatgttggggatttgctttag	336
8	132	agtcatcactgcccccaccctac**ag**	GAGACC...AAACAG	**gt**aagggtgtaagggtccttggc	590
9	209	gtatttttaccttctcacttccc**ag**	GAGAGC...GGCAGG	**gt**aagtggaatcaggccccaac	2351
10	175	cttctgttgtcctcgcattcccc**ag**	TATGAC...GCGATG	**gt**gaggccataagaatagtaaga	385
11	127	ctttctcctaacttgggtatacc**ag**	CTATGC...GAAAAA	**gt**gagcattctcataactcttcct	778
12	151	tggaattcactgtgtgtggctcc**ag**	GGAGAA...GCCAGG	**gt**gagtgcctgggggagcctag	3074
13	141	agcaatcctccctttttcttccct**ag**	ACATCT...AAGAAG	**gt**gaggattcagatgcctgggc	1117
14	104	ccagccacacttgccttggtctt**ag**	GTTCTG...GAAGAG	**gt**aaggtggggcaagcaaatga	1168
15	86	ctcttaactgctgtgggtattttt**ag**	AAGGAA...CTCAAG	**gt**gagtgagcccacttactgatg	466
16	70	ctctacaaccatcaaccctctga**ag**	TACACA...ATCCAG	**gt**gagtacggatgttgctcctgcc	1288
17	199[b]	tccccctttcttctgtctggggc**ag**	GTTGAA...ATCTGA		

[a]starting from the translation start site ATG. [b]ending in the stop codon TGA.

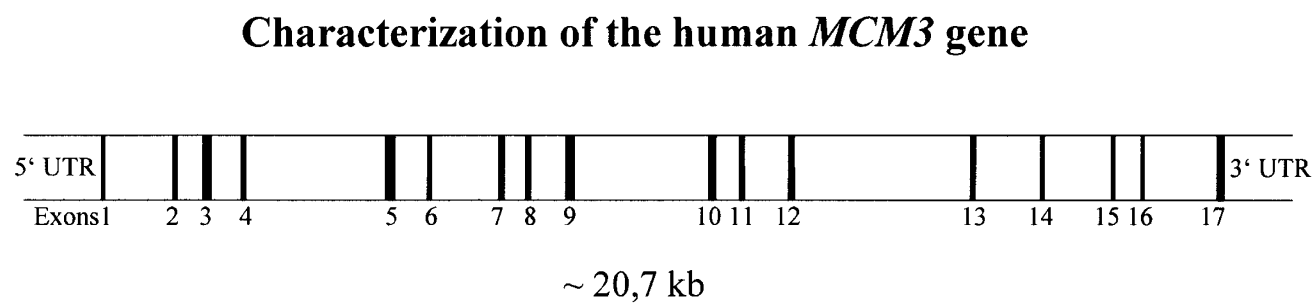

Figure 1 Characterization of the human MCM3 gene. Genomic organization of the human MCM3 gene. Exons are represented by black lines and introns by white boxes.

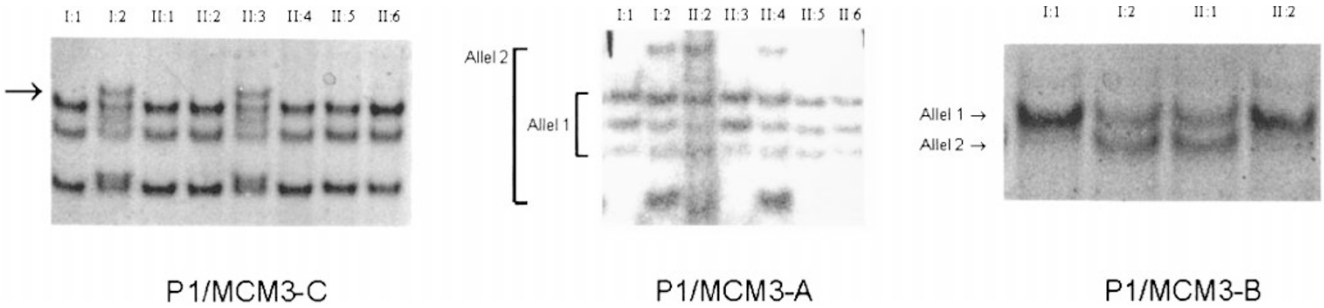

Figure 2 **a**) P1/MCM3–C, a polymorphism found in intron 2: two patterns of bands were identified; individuals I:1, II:1, II:2, II:4, II:5 and II:6 present only the allele-1 related pattern, whilst individuals I:2 and II:3 show patterns corresponding to alleles 1 and 2. The arrow indicates the allele 2 related bandshift in lanes 2 and 5. **b**) and **c**) P1/MCM3–A/B detected bandshifts related to the polymorphic site in bp No. 2329 of the coding region. Since both P1/MCM3–A and P1/MCM3–B are due to the same bp exchange in overlapping fragments of MCM3, the polymorphic marker was named P1/MCM3–A/B. For P1/MCM3–A, lanes 1, 4, 6 and 7 show only allele-1 related patterns, whilst in lanes 2, 3 and 5 the allele-2 related bandshift is shown. For P1/MCM3–B, individuals I:1 and II:2 present allele-1 related pattern, whilst individuals I:2 and II:1 show both allele 1 and allele-2 related bands.

families are identified with other recombinations mapping into the interval. These markers are also available now for prenatal diagnosis in ARPKD. Lastly, they may also be used to evaluate the gene as candidate for other disorders that map to chromosome 6p12.

Although we cannot exclude the possibility of mutations in the promotor or other regulatory regions within *MCM3* and that the sensitivity for mutation detection by SSCP is not complete, it is unlikely that pathogenic mutations would be restricted to the regulatory regions or that all mutations positioned within the coding region would be missed in our study. Our work, therefore, strongly suggests the exclusion of MCM3 as a candidate *PKHD1* gene.

Patient material and cDNA generation

Genomic DNA was isolated according to Miller *et al*.[11] Total RNA was isolated from 23 ARPKD-patients using Trizol (Life Technologies, Inc). The source for extraction included liver (10) and kidney samples (4), cultured fibroblasts (4), and EBV-transformed lymphoblasts (5). Additionally, using the same method, we isolated total RNA from seven controls (fetal liver (1), adult liver (1), fetal kidney (1), adult kidney (1), fresh blood samples (3)). The individual RNA samples were reverse-transcribed using oligo-dt primers, leading to first-strand cDNA synthesis. *MCM3*-specific cDNA products were then obtained by amplification with primer pairs covering the 2424 bp coding region and 34 and 74 bp, respectively, of adjacent 5' and 3' UTR sequences. In order to evaluate *MCM3* by SSCP, the *MCM3*-related cDNA was used to amplify 18 smaller fragments of the gene using primers designed in adjacent and overlapping pieces (primers available upon request). PCR products were analyzed by agarose gel electrophoresis and by SSCP. SSCP gels of 10% PAA (Acrylamide/Bisacrylamide 49:1) were run for 11–16 h both at room temperature and at 4°C for each fragment.

Acknowledgements

We thank the ARPKD families for supplying the blood and tissue samples. This study was supported by the Deutsche Forschungsgemeinschaft DFG and the National Institutes of Health under grant NIH RO1DK 51,259 (GGG, SS, LG W). This paper is a collaborative study by the autosomal recessive polycystic kidney disease consortium. Database accession numbers are GDB: x2153, d38073; OMIM: 263200

References

1 Zerres K, Muecher G, Bachner L *et al*: Mapping of the gene for autosomal recessive polycystic disease (ARPKD) to chromosome 6p21–cen. *Nat Genet* 1998; **7**: 429–432.
2 Guay-Woodford LM, Galliani CA, Musulman-Mroczek E, Spear GS, Guillot AP, Bernstein J: Diffuse renal cystic disease in children: morphologic and genetic correlations. *Pediatr Nephrol* 1998; **12**: 173–182.
3 Guay-Woodford LM: Autosomal recessive disease: clinical and genetic profiles. In: Torres V, Watson M, *Polycystic Kidney Disease*. Oxford University Press: Oxford, 1996; 237–267.
4 Zerres K, Muecher G, Becker J *et al*: Prenatal diagnosis of autosomal recessive polycystic kidney disease (ARPKD): molecular genetics, clinical experience, and fetal morphology. *Am J Med Genet* 1998; **76**: 137–144.
5 Muecher G, Wirth B, Zerres K: Refining the map and defining flanking markers of the gene for autosomal recessive polycystic kidney disease on chromosome 6p21.1–p12. *Am J Hum Genet* 1994; **55**: 1281–1284.
6 Muecher G, Becker J, Knapp M *et al*: Fine mapping of the autosomal recessive polycystic kidney disease locus (PKHD1) and the genes MUT, RDS, CSNK2B and GSTA1 at 6p21.1–p12. *Genomics* 1998; **48**: 40–45.
7 Thoemmes P, Fett R, Schray B *et al*: Properties of the nuclear P1 protein, a mammalian homologue of the yeast Mcm3 replication protein. *Nucleic Acids Res* 1992; **20**: 1069–1074.
8 Lens X, Onuchic L, Wu G *et al*: An integrated genetic and physical map of the autosomal recessive polycystic kidney disease region. *Genomics* 1997; **41**: 463–466.
9 Park J, Dexit M, Onuchic L *et al*: A 1 Mb BAC/PAC-based physical map of the autosomal recessive polycystic kidney disease gene (PKHD1) region on chromosome 6. *Genomics* 1999; **57**: 249–255.
10 Kubota Y, Mimura S, Nishimoto S, Takisawa H, Nojima H: Identification of the yeast MCM3-related protein as a component of Xenopus DNA replication licensing factor. *Cell* 1995; **81**: 601–609.
11 Miller SA, Dykes DD, Polesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988; **16**: 1215.