Evidence for linkage disequilibrium between \textit{D16S94} and the adult onset polycystic kidney disease (PKD1) gene

S E Pound, A D Carothers, P M Pignatelli, A M Macnicol, M L Watson, A F Wright

Genetic mapping of the PKD1 gene in chromosomal region 16p13.3 has depended on the identification of closely linked markers and rare recombinants that suggest the order of loci in relation to meiotic crossover points.\textsuperscript{1,2} This has led to the proposal that the PKD1 gene lies in an approximately 750 kb region close to marker loci \textit{D16S84}, \textit{D16S125}, and \textit{D16S94}, and proximal to the hypervariable probe \textit{D16S85} that first detected linkage to PKD1.\textsuperscript{3} However, genetic (meiotic) mapping can sometimes lead to inconsistent localisations, as with the Huntington’s disease gene,\textsuperscript{4} owing to factors such as rare double crossovers, gene conversion events, gonadal mosaicism, genetic heterogeneity, misdiagnosis, mistyping, and non-paternity. While many of these factors can be minimised or excluded, some are very difficult to exclude and, although rare, become significant compared with the frequency of rare recombination events separating markers from disease genes. In the absence of patients with chromosomal rearrangements that can help to localise the PKD1 gene physically, genetic methods of mapping are critical.

A means of genetically mapping a disease gene, other than by linkage analysis, is by the identification of non-random allelic association (linkage disequilibrium) between a marker and disease locus. This can occur if a mutation causing disease arises in close association with a relatively uncommon marker allele or haplotype, which can then become over-represented in the patient group relative to the general population. The finding of linkage disequilibrium has served to refine the mapping both of the cystic fibrosis gene\textsuperscript{5} and of the Huntington’s disease gene.\textsuperscript{6} However, linkage disequilibrium is expected to occur only if the mutation rate is relatively low or if the frequency of a particular chromosomal haplotype is inflated in the population of affected subjects as a result of previous selective advantage, migration, or population stratification. It requires that many or most affected subjects carry the same original mutation.

It has been suggested that the mutation rate in ADPKD is high,\textsuperscript{7} based on the apparently high frequency of unaffected parents of ADPKD subjects and on the high population prevalence. This study was carried out at a time when methods of presymptomatic diagnosis were less sensitive than with ultrasonography and when paternity testing was less readily available. We therefore tested a genetically homogeneous set of 33 ADPKD families for evidence of linkage disequilibrium between the PKD1 locus and closely linked markers \textit{D16S84}, \textit{D16S125}, and \textit{D16S94}.

A set of 34 ADPKD families was ascertained through medical renal units in Edinburgh and Glasgow and the diagnoses established as described.\textsuperscript{8,9} Family members were typed for the markers \textit{D16S85} (3’HVR), \textit{D16S84} (CMM65), \textit{D16S125} (26-6PROX), and \textit{D16S94} (VK5). Multipoint heterogeneity analysis was carried out to determine the proportion of families unlinked to the PKD1 region of chromosome 16.\textsuperscript{10} The sample was found to show evidence of linkage heterogeneity but 90 to 95% of the kindreds were linked to \textit{PKD1}. A single unequivocally unlinked family was removed from the analysis, leaving 33 families for genetic mapping of the PKD1 locus.\textsuperscript{11} In 28 families the haplotypes could be determined unambiguously, in one family the haplotype could not be established, and in the remaining four this was done by minimising the number of crossovers. The normal chromosomes used to establish control allele frequencies were derived from the unaffected spouses of family members.

The results are shown in the table. No evidence of allelic association was found between \textit{PKD1} and either \textit{D16S84} or \textit{D16S125}, both of which show close genetic linkage to \textit{PKD1}.\textsuperscript{12-14} In the case of \textit{D16S94}, there is evidence of linkage disequilibrium, with the affected haplotypes in seven kindreds showing the 1·6 kb allele (allele 1) and in 22 kindreds the 1·3 kb allele (allele 2) compared with the control population, who showed 38 1·6 kb alleles and 34 1·3 kb alleles. This result is significant at the 2% level (Yates’s corrected $\chi^2=5·75$). The allele frequencies in the control population are in good agreement with those published previously.\textsuperscript{15}

The suggestion of linkage disequilibrium between \textit{PKD1} and \textit{D16S94} has implications for the genetic localisation of \textit{PKD1}. Any initial disequilibrium occurring when a mutation arises is dissipated at the rate of (1-r)$^n$ in n generations, where r is the recombination
fraction separating marker and disease locus. If \( r \) is very small, the rate of approach to equilibrium is slow. Similarly, a recent mutation will take a finite time to reach equilibrium with adjacent markers, even when \( r \) is significant. In practice, a plateau of linkage disequilibrium can define an area of up to several hundred kilobases of DNA on each side of the gene of interest, in the absence of recombinational hot spots. The identification of meiotic crossover points by haplotype analysis has led to the proposal that PKD1 lies between D16S84 and D16S125, a region shown by pulsed field electrophoresis to contain less than 750 kb of DNA and found to contain a large number of genes.4 The more proximally located marker D16S94 lies just outside this area. If the presence of linkage disequilibrium can be confirmed by other groups with ethnically homogeneous patient populations, it would argue for a detailed physical mapping effort in the proximal part of the D16S84-D16S125 interval, although further marker studies are required to establish the extent of the region of disequilibrium. The failure to find any disequilibrium with D16S125, located about 50 to 100 kb distally, could be because a common PKD1 mutation arose in association with the more frequent D16S125 allele, making it more difficult to detect disequilibrium, or because the TaqI polymorphism is itself subject to significant mutational change.

The high population prevalence of ADPKD (1 in 1000), 90 to 95% of which are the result of PKD1 gene mutations,4,10 is consistent with a high mutation rate but since the reproductive fitness of ADPKD patients is also high, this is not necessarily so. Additional factors, such as reproductive compensation, previous selective advantage, and genetic drift, may also contribute to the high prevalence. Linkage disequilibrium is not expected to occur if the mutation rate is high since each new mutation is likely to occur on a different chromosomal background.

The results described in these ADPKD families therefore add support to the idea that the high prevalence is maintained by a high reproductive fitness or other factors referred to above, in the presence of an intermediate or low mutation rate. These issues await further clarification when the gene is isolated.

The financial support of The Wellcome Trust is gratefully acknowledged. We would also like to thank Drs J D Briggs and B Junor for access to their patients. SEP is in receipt of an MRC Training Fellowship.