Am. J. Hum. Genet. 56:1315–1319, 1995

Haplotypes and Mutations in Wilson Disease

Gordon R. Thomas,[1,2] Eve A. Roberts,[1,3] John M. Walshe,[4] and Diane W. Cox[1,2,3]

[1]Research Institute, The Hospital for Sick Children, and Departments of [2]Molecular and Medical Genetics and [3]Paediatrics, University of Toronto, Toronto; and [4]Middlesex Hospital, London

Summary

Wilson disease is a disorder of copper transport, resulting in neurological and hepatic damage due to copper toxicity. We have recently identified >20 mutations in the copper-transporting ATPase defective in this disease. Given the difficulties of searching for mutations in a gene spanning >80 kb of genomic DNA, haplotype data are important as a guide to mutation detection. Here we examine the haplotypes associated with specific mutations. We have extended previous studies of DNA haplotypes of dinucleotide-repeat polymorphisms (CA repeats) in the Wilson disease region to include an additional marker, in 58 families. These haplotypes, combining three markers (D13S314, D13S316, and D13S301), are usually specific for each different mutation, even though highly polymorphic CA repeat markers have been used. Haplotypes, as well as their accompanying mutations, differ between populations. In the patients whom we have studied, the haplotype data indicate that as many as 20 mutations may still be unidentified. The use of the haplotypes that we have identified provides an important guide for the identification of known mutations and can facilitate future mutation searches.

Introduction

Wilson disease is an autosomal recessive disorder of copper transport and has a worldwide frequency of between 1/5,000 and 1/30,000 live births (Danks 1989; Bonné-Tamir et al. 1990; Houwen et al. 1993*b*). The disorder manifests as chronic liver disease and/or neurological impairment due to accumulation of copper in several tissues, principally the liver and brain. Sudden death, from either hemolytic crisis or liver failure, or lifelong neurological disability can occur when the disease is identified too late.

The gene defective in Wilson disease has been identified and encodes a putative copper-transporting P-type ATPase (*ATP7B*) (Bull et al. 1993; Tanzi et al. 1993). A total of 25 disease-causing mutations have recently been identified in this gene (Bull et al. 1993; Tanzi et al. 1993; Thomas et al. 1995*a*). We have shown that the specific mutation appears to influence phenotype (Thomas et al. 1995*a*). We previously identified haplotypes specific to WND chromosomes in 51 patients from Canada and the United Kingdom and estimated the existence of at least 10 mutations in our northern European patients (Thomas et al. 1994). We have extended our haplotype analysis in a total of 58 families, with the addition of another CA repeat marker, and we have examined the relation between these haplotypes and the mutations identified in the WND gene in these families. Each haplotype is generally associated with a specific mutation, indicating that the haplotype data will be important for guiding the identification of mutations, both established and currently unidentified.

Patients and Methods

Patients

DNA was isolated from peripheral blood (Miller et al. 1988) collected from 32 Canadian families (23 of northern European, 4 of Indian, 3 of Mediterranean, and 2 of Oriental origin), 25 families from the United Kingdom (11 of northern European, 8 of Mediterranean, 5 of Indian, and 1 of Middle Eastern origin), and 1 family from Saudi Arabia. Haplotypes for 49 of these families were obtained by using D13S133, D13S314, and D13S316 in a previous study of 51 families. The remaining two families were not included, because of lack of additional DNA sample to carry out further analysis. The 58 families used in the current study consisted of 99 parents, 87 patients, and 82 unaffected sibs. Within each family, parents and patients were typed to determine the marker haplotypes present on the WND and normal chromosomes. Complete haplotypes were reconstructed for 114 of 116 WND chromosomes and 97 of 116 normal chromosomes (the lack of sample from one of the parents in some families made it impossible to determine the haplotypes of all 116 normal chromosomes). The age at onset of the patients ranged from 3 to 27 years: 64% of the patients had an age at onset of <16 years (74% of these had hepatic disease), and 36%

Received November 15, 1994; accepted for publication March 3, 1995.

Address for correspondence and reprints: Dr. D. W. Cox, Department of Genetics, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8. E-mail: dcox sickkids.on.ca

0002-9297/95/5606-0008$02.00

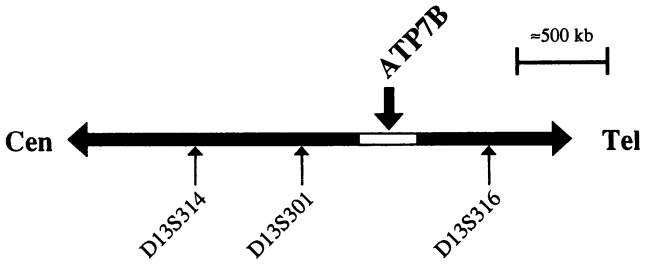

Figure 1 Diagram of the *ATP7B* region. The relative locations of the markers in this study are shown. Positions of the markers have been determined by further analysis of recombinant families (Thomas et al. 1993) and by analysis of overlapping YAC clones (Thomas et al. 1994).

had an age at onset ≥16 years (26% of these had liver disease). The ethnic origins of the parents of each patient were determined, where possible. Diagnosis of Wilson disease was originally established by biochemical assays as described elsewhere (Houwen et al. 1993*a*).

Haplotype Analysis

Typing of CA repeats was carried out in patients and their parents. Haplotypes of the markers D13S133, D13S314, D13S315, and D13S316 have been described elsewhere (Thomas et al. 1994). In order to include the additional marker D13S301 (Petrukhin et al. 1993) in our haplotypes, we determined its location relative to recombinant families described elsewhere (Thomas et al. 1993). The use of marker D13S133 was discontinued because of the presence of an allele that is very common on both normal and WND chromosomes (allele 17) and because it did not add information to the haplotypes. The locations of D13S314 and D13S316 have been described elsewhere (Thomas et al. 1994).

The amplification of CA repeats in patient and parent DNA were as described by Thomas et al. (1994). Primers used for D13S301 were as follows: 5′-ATCATACCTGGTTGTGCAA-3′ and 5′-TGATGCTTCTTTCTAAACACA-3′. Allele numbers were defined on the basis of CEPH parent 1332-02, who has alleles 5 (144 bp) and 8 (138 bp).

Results

The marker D13S301 was found to recombine with WND in family W41, placing it centromeric to the gene. The locations of the three markers used are shown in figure 1.

Haplotypes of the markers D13S301, D13S314, and D13S316 were constructed, and these data are presented in table 1. The names of the new haplotypes that include D13S301 are based on the nomenclature previously used by Thomas et al. (1994). Subgroups were determined by combining haplotypes that differ by ≤2 bp at a single locus, as had been done for the haplotypes described elsewhere (Thomas et al. 1994). A total of 15 of the 25 WND haplotypes in northern Europeans were not found on normal chromosomes.

WND haplotypes that occur more than once in populations other than northern Europeans have also been identified. These haplotypes are indicated in table 2. Some of the haplotypes are identical to those found in northern Europeans and are named accordingly.

To date, a total of 25 mutations in *ATP7B* have been described (Bull et al. 1993; Tanzi et al. 1993; Thomas et al. 1995*a*), and 23 of these are present in the families studied. All haplotypes have been analyzed for the presence of all known mutations. The mutations identified on the chromosomes are indicated for each haplotype in tables 1 and 2. Two mutations, N1271S (Tanzi et al. 1993) and R1320X (Thomas et al. 1995*a*), were found on chromosomes for which complete haplotypes could not be determined, and they are not included in table 1.

Discussion

The addition of one marker to the previously defined haplotypes has increased the number of different haplotypes. The two haplotype groups described elsewhere (Thomas et al. 1994) as being the most common in the northern European populations (i.e., haplotype groups A and B) have been shown to be a collection of different subhaplotypes when the additional marker is added, and the subgroups are generally associated with different mutations. The large number of subgroups may be a result of the increased variability of CA repeat markers. However, haplotypes C–E remain tightly grouped, with only slight variation (≤2 bp at the D13S301 locus), indicating that the various haplotypes in groups A and B likely represent multiple mutations.

We have found that the correlation between specific mutation and specific haplotype is high. The group D haplotype is associated exclusively with G1267R, and several other haplotypes (H and J–N) carry only a single mutation not found on any other chromosomes. Additionally, in most cases where mutations have been found on unique WND haplotypes, these mutations were not observed in any other patient. Thus, CA repeat markers close to a disease locus are sufficiently stable to be used in linkage disequilibrium studies, as has been shown for several other disorders (McClatchey et al. 1992; Sirugo et al. 1992; Oudet et al. 1993).

There are some exceptions to the correlation between mutations and haplotypes. Groups C and E (which differ by 4 bp at D13S314) are likely to share a common origin, since the majority of both haplotypes carry the H1070Q mutation. The variation between the group C and E haplotypes likely represents an allele slippage event at the D13S314 marker, which resulted in a vari-

Table 1

Haplotype Distribution on Chromosomes of Northern European Origin

Group	Original Haplotype[a]	Subgroup	Extended Haplotype[b]	Mutation	WND	Normal
A	10-7	A1	10-27		2	0
			10-37		1	1
		A2	10-47	2578+1G→C[c]	1	0
			10-47		3	1
		A3	10-57		2	1
			10-67		2	0
		A4	10-77	3651del6[c]	1	0
		A5	10-177	1747delAT[c]	1	0
	10-8	A6	10-38	2302insC[c]	1	0
B	11-7	B1	11-37		2	0
			11-47		2	1
		B2	11-67		2	4
			11-77		1	1
C	10-6	n/a	10-46	H1070Q[d]	14	0
			10-46		1	0
			10-56	H1070Q[d]	2	0
			10-56		1	0
D	3-5	n/a	3-35	G1267R[c]	7	0
E	11-6	n/a	11-46	H1070Q[d]	3	0
			11-46		2	0
F	3-7	F1	3-47	3559+1G→A[c]	1	0
	4-8	F2	4-48		1	0
G	7-2	n/a	7-72	3088delAC[c]	1	0
			7-73		1	0
None	n/a	M	11-74	3629del4[c]	1	0
		...	12-47	2302insC[c]	1	3
		...	3-34	3403delC[d]	1	1
		...	12-67	4094delTG[c]	1	1
		...	8-51		1	0
		...	13-57		1	0
		...	11-85		1	0
		...	1-67		1	0
		...	1-34		1	2
		...	7-84		1	2
		...	6-53		1	0
		21 others			0	32
Total					66	50

NOTE.—n/a = not applicable.

[a] As defined elsewhere (Thomas et al. 1994), given in the order D13S314–D13S316. Note that alleles 10–13 of D13S314 are 4 bp apart.

[b] Extended versions of original haplotypes, given in the order D13S314–D13S301–D13S316.

[c] Thomas et al. (1995*a*).

[d] Tanzi et al. (1993).

ant haplotype. The mutation could be old, allowing time for this haplotype variation. This mutation has also been identified in the Mediterranean group, which is not clearly separable from the northern European group.

Two mutations appear to be present on very different haplotypes (i.e., haplotypes that differ by >2 bp at a single marker). One source of such variation could be recombination with normal haplotypes, which we have directly observed for a mutation in the Icelandic population (Thomas et al. 1995*b*). The R779L mutation is present on three haplotypes in our Chinese patients and is unlikely to represent independent occurrences of the same mutation. The mutation may be very old, and multiple recombination events may have resulted in the variant haplotypes. The 2302insC mutation is also present on three very different haplotypes. This mutation is an insertion of a C into a series of six C's within the cDNA sequence and may represent a spot where DNA polymerase is more prone to replication errors. Therefore, these three haplotypes may represent independent origins of the same mutation.

Another exception to the haplotype/mutation correla-

Table 2

WND Haplotypes on Non–Northern European Chromosomes

Ethnic Origin and Group	Haplotype[a]	Mutation	WND	Normal
Indian:				
H	7-6-4	I1103T[b]	2/16	0/16
	7-7-4		1/16	0/16
I	5-3-4		2/16	0/16
J	3-7-4	3149delC[b]	2/16	0/16
K	10-17-4	G1102R[b]	2/16	0/16
L	12-6-7	1711-1G→C[b]	2/16	0/16
B1	11-4-7		1/16	0/16
	11-4-8		1/16	1/16
B2	11-8-7		1/16	0/16
M	11-7-4		1/16	0/16
...	3-4-4	1748insT[b]	1/16	0/16
...	5-6-4		1/16	1/16
Bangladeshi:				
B1	11-3-7	G944S[b]	1/2	0/2
	11-4-7	G944S[b]	1/2	0/2
Middle Eastern:				
M	11-7-4	L937X[b]	1/4	0/4
	11-7-5	L937X[b]	1/4	0/4
N	11-4-4	846delC[b]	1/4	0/4
...	6-7-6		1/4	0/4
Chinese:				
...	5-8-5.5	R779L[b]	1/4	0/3
...	6-4-7	R779L[b]	1/4	0/3
...	6-8-7	R779L[b]	1/4	0/3
...	4-4-7		1/4	0/3
Mediterranean:				
A3	10-5-7		1/22	0/22
	10-6-7		1/22	0/22
B1	11-4-7		4/22	1/22
C[c]	10-4-6	H1070Q[d]	2/22	0/22
G	7-7-3		1/22	1/22
N	10-4-4	846delC[b]	2/22	1/22
...	11-8-5	L937X[b]	1/22	0/22
...	1-3-8	2161delA[b]	1/22	0/22
...	10-17-5	2302insC[b]	1/22	0/22
...	15-4-7		1/22	0/22
...	7-4-4		1/22	1/22
...	9-7-4		1/22	1/22

[a] In the order D13S314–D13S301–D13S316.
[b] Thomas et al. (1995*a*).
[c] Same haplotype with the same mutation occurs in the northern European population.
[d] Tanzi et al. (1993).

tion is the occurrence of more than one mutation on a single haplotype. Some cases of this are likely to be different mutation events occurring on the same common haplotype background (e.g., 2578+1→C on one of the four group A2 haplotypes). Other cases may represent population differences in mutation spectrum (e.g., the B1 haplotype in the Bangladeshi patient carries G944S, whereas B1 haplotypes from other ethnic groups do not). However, four of the C/E haplotypes do not carry the H1070Q and must have another mutation, which has not been identified. This haplotype was not found on 54 normal chromosomes and is not likely to represent a common background on which mutations have occurred. Identification of the mutation(s) on these haplotypes may shed some light on this question.

We have revised our estimate of the number of mutations responsible for Wilson disease, to include the new marker and haplotype data. Assuming that 2-bp variations are more likely to represent CA repeat slippage than different mutation events, we estimate that there may be on the order of 16 additional mutations in our northern European patients and 9 additional mutations in the other ethnic groups. The 23 mutations present in our patients represent ~48% (23 of 48) of the haplotypes present on these chromosomes. The number of mutations in Wilson disease may reflect both the important role of this protein in protection from copper toxicity and the many ways its function can be impaired.

The D13S316 marker can specifically point to the presence of the two most common mutations. Allele 6 at the marker is found on most chromosomes with H1070Q, while allele 5 is exclusively associated with G1267R. The use of this single marker will identify ~40% (26 of 66) of the mutations in northern European chromosomes, greatly reducing the amount of screening necessary.

Mutation detection in Wilson disease is challenging because of the presence of a large number of mutations in a 4.1-kb coding region in 22 exons spread over >80 kb of genomic DNA. The haplotype data presented here will be of great value in mutation detection, since identification of haplotypes for which a mutation has been found would allow confirmation of the mutation by direct analysis where available (Thomas et al. 1995*a*, 1995*b*) or by direct sequencing. The identification of mutations, as well as the role of haplotypes in this search, is important as a guide to prognosis and perhaps even to treatment. We have recently shown that mutations that destroy the ATPase are associated with severe liver disease in early childhood, at an age when Wilson disease has not usually been suspected (Thomas et al. 1995*a*).

Acknowledgments

We thank the families for their cooperation and understanding; Drs. A. N. Bekassy, R. Harris, M. Lambert, S. Moroz, H. Nazer, T. Rosales, C. Seymour, K. Westermark, and R. Williams for providing patient samples; Dr. J. Ault for locating many of the stored U.K. DNA samples; and R. F. Wintle for critical review of the manuscript. This study was supported by the Networks of Centres of Excellence, Canadian Genetic Diseases Network. G.R.T. was supported by scholarships from the Canadian Liver Foundation and the Medical Research Council of Canada.

References

Bonné-Tamir B, Frydman M, Agger MS, Bekeer R, Bowcock AM, Hebert JM, Cavalli-Sforza LL, et al (1990) Wilson's disease in Israel: a genetic and epidemiological study. Ann Hum Genet 54:155–168

Bull PC, Thomas GR, Rommens JM, Forbes JR, Cox DW (1993) The Wilson disease gene is a putative copper transporting P-type ATPase similar to the Menkes gene. Nat Genet 5:327–337

Danks DM (1989) Disorders of copper transport. In: Beaudet AL, Sly WS, Valle D (eds) The metabolic basis of inherited disease, 6th ed. McGraw-Hill, New York, pp 1411–1431

Houwen RHJ, Thomas GR, Roberts EA, Cox DW (1993*a*) DNA markers for the diagnosis of Wilson disease. J Hepatol 17:269–276

Houwen RHJ, van Hattum J, Hoogenraad TU (1993*b*) Wilson disease. Neth J Med 43:26–37

McClatchey AI, Trofatter J, McKenna-Yasek D, Raskind W, Bird T, Pericak-Vance M, Gilchrist J, et al (1992) Dinucleotide repeat polymorphisms at the *SCN4A* locus suggest allelic heterogeneity of hyperkalemic periodic paralysis and paramyotonia congenita. Am J Hum Genet 50:896–901

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215

Oudet C, Mornet E, Serre JL, Thomas F, Lentes-Zengerling S, Kretz C, Deluchat C, et al (1993) Linkage disequilibrium between the fragile X mutation and two closely linked CA repeats suggests that fragile X chromosomes are derived from a small number of founder chromosomes. Am J Hum Genet 52:297–304

Petrukhin KE, Fischer SG, Pirastu M, Tanzi RE, Chernov I, Devoto M, Brzustowicz LM, et al (1993) Mapping, cloning and genetic characterization of the region containing the Wilson disease gene. Nat Genet 5:338–343

Sirugo G, Keats B, Fujita R, Duclos F, Purohit K, Koenig M, Mandel JL (1992) Friedreich ataxia in Louisiana Acadians: demonstration of a founder effect by analysis of microsatellite-generated extended haplotypes. Am J Hum Genet 50:559–566

Tanzi RE, Petrukhin KE, Chernov I, Pellequer JL, Wasco W, Ross B, Romano DM, et al (1993) The Wilson disease gene is a copper transporting ATPase with homology to the Menkes disease gene. Nat Genet 5:344–350

Thomas GR, Bull PC, Roberts EA, Walshe JM, Cox DW (1994) Haplotype studies in Wilson disease. Am J Hum Genet 54:71–78

Thomas GR, Forbes JR, Roberts EA, Walshe JM, Cox DW (1995*a*) The Wilson disease gene: the spectrum of mutations and their consequences. Nat Genet 9:210–217

Thomas GR, Jensson O, Gudmundsson G, Thorsteinsson L, Cox DW (1995*b*) Wilson disease in Iceland: a clinical and genetic study. Am J Hum Genet 56:1140–1146

Thomas GR, Roberts EA, Rosales TO, Moroz SP, Lambert MA, Wong LTK, Cox DW (1993) Allelic association and linkage studies in Wilson disease. Hum Mol Genet 2:1401–1405