Complex inheritance of familial hypercholanemia with associated mutations in *TJP2* and *BAAT*

Victoria E. H. Carlton[1], Baruch Z. Harris[2], Erik G. Puffenberger[3], A. K. Batta[4], A. S. Knisely[5], Donna L. Robinson[3], Kevin A. Strauss[3], Benjamin L. Shneider[6], Wendell A. Lim[7], Gerald Salen[8], D. Holmes Morton[3] & Laura N. Bull[1]

Published online 21 April 2003; doi:10.1038/ng1147

Familial hypercholanemia (FHC) is characterized by elevated serum bile acid concentrations, itching, and fat malabsorption[1,2]. We show here that FHC in Amish individuals is associated with mutations in tight junction protein 2 (encoded by *TJP2*, also known as *ZO-2*) and bile acid Coenzyme A: amino acid N-acyltransferase (encoded by *BAAT*). The mutation of TJP2, which occurs in the first PDZ domain, reduces domain stability and ligand binding *in vitro*. We noted a morphological change in hepatic tight junctions. The mutation of BAAT, a bile acid–conjugating enzyme[3], abrogates enzyme activity; serum of individuals homozygous with respect to this mutation contains only unconjugated bile acids. Mutations in both *TJP2* and *BAAT* may disrupt bile acid transport and circulation. Inheritance seems to be oligogenic, with genotype at *BAAT* modifying penetrance in individuals homozygous with respect to the mutation in *TJP2*.

We have identified 17 individuals with FHC in 12 families of Lancaster County Old Order Amish descent (Fig. 1; Table 1). Serum bile acid concentration in affected individuals fluctuated (often very high, occasionally normal). Fat malabsorption, reflecting low intestinal bile acid levels, was manifested by failure to thrive, potentially life-threatening vitamin-K dependent coagulopathy and rickets. Symptoms usually responded to treatment with ursodeoxycholic acid (UDCA). FHC is atypical for a liver disease: test results of biochemical markers of liver injury were normal, except for alkaline phosphatase activity, which sometimes rose. Liver biopsy findings varied. One untreated individual (1d) had canalicular cholestasis and two individuals (7c, 12c) receiving UDCA had minimally active chronic hepatitis[1]. Several older individuals have become symptom-free and have discontinued UDCA treatment. We have no information on the natural history of FHC in adults.

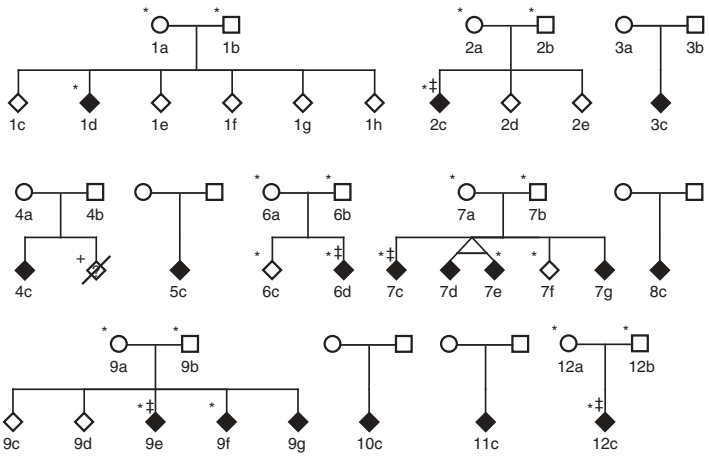

Fig. 1 Families with FHC. All families are of Lancaster County Old Order Amish descent. Unaffected siblings who were genotyped are shown. The child marked with a plus sign (+) was not genotyped but died of vitamin K–responsive coagulopathy at 4 months and suffered from itching, a phenotype consistent with FHC. Individuals marked with an asterisk (*) were included in the genome screen; those marked with a double dagger (‡) were included in the initial analysis of genome screen data.

A whole-genome screen identified one chromosomal region (9q12–q13) shared identically by descent (IBD) on most (6 of 10) chromosomes of affected individuals included in the initial analysis. Several 9q12–q13 markers were in linkage disequilibrium with FHC (Fig. 2*a*). We examined the region in additional individuals with FHC. Of 12 individuals genotyped, 8 shared the haplotype in 9q12–q13 (Fig. 2*b*), a region containing the candidate gene *TJP2*. Genomic sequencing of exons and exon-intron boundaries of *TJP2* in individual 3c identified a mutation, 143T→C, predicted to cause a valine-to-alanine substitution (V48A) in TJP2. Screening of all 17 individuals with FHC identified 11 individuals (including a pair of monozygotic twins) in 8 families who were homozygous with respect to the mutation ($TJP2^{143C/143C}$; Table 1). Three unaffected siblings were also homozygous with respect to the mutation, showing that penentrance is incomplete. The mutation was not present in 190 control chromosomes from Caucasian individuals, 34 control chromosomes from Amish (non-Lancaster County) individuals or 15 non-transmitted parental chromosomes. It was seen on 7 of 104 control chromosomes from Lancaster County Old Order Amish individuals.

The mutated valine is in the N-terminal PDZ domain (PDZ1) of TJP2 and is highly conserved in the domain family (it is invariably a

[1]*Liver Center Laboratory and Department of Medicine, San Francisco General Hospital, University of California San Francisco, California 94110, USA.* [2]*Program in Biological Sciences, University of California San Francisco, California 94143, USA.* [3]*Clinic for Special Children, Strasburg, Pennsylvania 17579, USA.* [4]*Veterans Affairs Medical Center, East Orange, New Jersey 07018, USA.* [5]*Institute of Liver Studies, King's College Hospital, Denmark Hill, London SE5 9RS, UK.* [6]*Division of Pediatric Hepatology, Department of Pediatrics, Mount Sinai School of Medicine, New York City, New York 10029, USA.* [7]*Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94143-0450, USA.* [8]*University of Medicine and Dentistry, New Jersey Medical School, Newark, New Jersey 07103, USA. Correspondence should be addressed to L.N.B .(e-mail: lbull@medsfgh.ucsf.edu).*

branched hydrophobic residue: valine, leucine or isoleucine; ref. 4). In all known PDZ structures, this residue is partially buried in the protein core and lies at the base of the peptide-binding groove[5], suggesting importance for domain stability and ligand binding. We investigated effects of the V48A substitution on TJP2-PDZ1.

Circular dichroism spectroscopy of wild-type and V48A TJP2-PDZ1 domains indicated that the mutation reduced domain stability. V48A had significantly less ellipticity at 222 nm (Fig. 3*a*), implying loss of α-helical structure, and did not exhibit co-operative unfolding transition upon chemical denaturation (Fig. 3*b*), indicating that mutant protein was largely unfolded, even in the absence of denaturant. Lower stability of V48A TJP2-PDZ1 is consistent with our observation that V48A TJP2-PDZ1 was poorly expressed in bacterial cells and was less soluble than wild-type TJP2-PDZ1 (data not shown).

We tested *in vitro* binding of TJP2-PDZ1 to five physiologically relevant known or postulated ligands (claudin C-terminal peptides[6] with known liver expression[7]). Fluorescence anisotropy indicated that the substitution variably lessened binding to all claudins tested (Fig. 3*c*,*d*). Changing relative binding affinities of TJP2 for different claudins could change ratios of claudins in tight junctions.

TJP2 is a ubiquitous tight junction scaffold protein present in liver[8]. Tight junctions are intercellular barriers that control paracellular solute diffusion across cell layers[9]. In liver, they separate bile from plasma. The physical barrier of tight junctions is composed of strands of claudins and occludin[10]. Humans express approximately 18 claudins. The complement of expressed claudins differs among tissues and affects junction permeability[11]. Through binding to claudin and occludin, TJP2 is presumed to affect tight junction structure and the open or closed state of hypothesized claudin pores[10]. In liver, tight junctions maintain a 200–10,000-fold bile acid concentration gradient[12]. We propose that the V48A substitution increases paracellular permeability to bile acids (and probably to other small biliary anions as well), which leak from bile into plasma, leading to lower biliary and intestinal bile acid concentrations. Possible mechanisms for greater permeability include the following: (i) changes in ratios of claudins in tight junctions, (ii) failure of claudin strands to anchor properly to tight junctions, leading to breaks, or (iii) changes in the open or closed state of claudin pores.

We noted a morphological abnormality of tight junctions in liver tissue from individual 7c. This was the only tissue suitable for ultrastructural (transmission electron microscopy) studies that was available from an individual with FHC. Individual 7c was homozygous with respect to the mutation in *TJP2* and heterozygous with respect to the mutation in *BAAT* (see below). Tight junction depth was significantly greater in individual 7c than in two control individuals (Table 2). A rodent model of liver disease, the bile duct–ligated rat, has greater junction permeability and depth[13,14]. Expanded tight junctions and aberrant strands are formed in cells overexpressing a mutant claudin 1 with impaired binding to TJP2 (and to TJP1 and TJP3). These cells also have greater paracellular permeability for some solutes[10]. Greater tight junction depth may thus reflect abnormalities in structure and permeability. We speculate that in individual 7c, claudin strands

Table 1 • Clinical and genotypic summary

						Symptoms at diagnosis					Current status		
Family	Individual	*TJP2*	*BAAT*	Affected	Maximum SBA[a] (g/ml)	Failure to thrive	Coagulopathy	Itching	Chronic URI	Jaundice	If symptomatic, current age	If asymptomatic, since what age	γ-GT[b] mean ± s.d. (*n*)
	Individuals with FHC												
1	1d	143C/143C	+/+	+	46[c]	+	+	+	–	+		6	29 ± 7 (5)
2	2c	143C/143C	+/+	+	105	–	+	+	+/–	–		5	24 ± 2 (6)
3	3c	143C/143C	+/+	+	41	+	+	+	–	+	2		17 ± 10 (2)
4	4c	143C/143C	+/+	+	nd	–	+	–	–	–		10	23 (1)
5	5c	143C/143C	+/+	+	39[c]	+	+	+	–	–	2		nd
6	6c	143C/143C	+/+	+	140	+	–	+	–	–	6		24 ± 19 (6)
7	7c	143C/143C	226G/+	+	89	+	+	+	–	–	8		44 ± 19 (15)
7	7d[d]	143C/143C	226G/+	+	39	–	–	–	+	–		4	23 ± 3 (2)
7	7e[d]	143C/143C	226G/+	+	30	–	–	–	+	–		4	28 (1)
7	7g	143C/143C	226G/+	+	217	+	+	+	–	–	2		34 ± 9 (2)
8	8c	143C/143C	226G/+	+	14[c]	+	+	+	–	–	<1		nd
9	9e	+/+	226G/226G	+	115	+	–	–	+	–	8		9 ± 4 (5)
9	9f	+/+	226G/226G	+	43	+	–	+/–	+	–	6		8 ± 4 (3)
9	9g	+/+	226G/226G	+	132	–	–	–	–	–		birth[e]	6 (1)
10	10c	+/+	226G/226G	+	32[c]	+	+/–	+	–	–	unknown		12 (1)
11	11c	143C/+	226G/226G	+	20[c]	+	+	–	–	–	<1		nd
12	12c	+/+	+/+	+	164	+	–	+	+	–	4		11 ± 1 (4)
	Non-penetrant siblings[f]												
1–6	–	143C/143C	+/+	–	always asymptomatic, currently age 7 with normal SBA								
1–6	–	143C/143C	+/+	–	always asymptomatic, currently age 9 with normal SBA								
1–6	–	143C/143C	+/+	–	always asymptomatic, currently age 11 with unknown SBA								

[a]Unless noted, value is maximum serum bile acid concentration as determined by GLC. [b]Because normal γ-glutamyltranspeptidase (γ-GT) levels in infants are not well defined, we used only measurement taken while child was over 1 year of age. [c]No GLC values were available, and given values were determined from an enzymatic assay. [d]Monozygotic twins. [e]Asymptomatic since birth but has elevated SBA. [f]All children are in families 1 through 6. To maintain anonymity, specific family numbers are not given. nd, not determined; URI, upper respiratory infection.

a

marker	λ	*P* value
D9S1853	0.00	–
D9S1817	0.00	–
D9S1874	0.57	0.0110
D9S2148	0.00	–
D9S1844	0.57	–
fhc-gata1	0.85	0.0059
fhc-ga1	0.63	0.0160
fhc-ca2	0.48	–
D9S1862	0.57	0.0042
D9S1777	0.00	–
D9S234	0.50	–
fhc-ca1	0.62	0.0172
D9S1879	0.00	–
fhc-ca3	0.62	0.0172
D9S1787	0.51	–
D9S1800	0.64	0.0096
fhc-ca6	0.67	–
D9S273	0.59	0.0432
fhc-ga2	0.50	–
D9S166	0.65	0.0054
D9S1799	0.57	0.0042
D9S301	0.57	–
fhc-ca8	0.00	–
fhc-ca9	1.00	0.0008
fhc-ca10	1.00	0.0008
D9S1806	0.42	–
fhc-ga4	0.42	–
D9S1822	0.00	–
D9S175	0.00	–
D9S1834	0.00	–
D9S1674	0.00	–

b

marker	cM	Mb	1d	1d	2c	2c	3c	3c	4c	4c	5c	5c	6c	6c	7c	7c	7e	7e
D9S1853	56.3	29.9	269	269	257	257	–	–	–	–	–	–	257	257	265	257	265	257
D9S1817	59.3	35.5	282	298	308	296	–	–	–	–	–	–	298	296	298	298	298	298
D9S1874	61.4	38.9	207	217	205	205	–	–	–	–	–	–	205	205	205	205	205	205
D9S2148	63.7	39.9	164	152	152	152	152	152	–	–	–	–	152	152	152	156	152	156
D9S1844	64.7	–	251	243	247	247	–	–	–	–	–	–	247	247	247	247	247	247
fhc-gata1	–	43.1	1	1	1	1	–	–	–	–	–	–	1	1	1	1	1	1
fhc-ga1	–	43.1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
fhc-ca2	–	43.1	2	2	2	2	–	–	–	–	–	–	2	2	2	2	2	2
D9S1862	64.7	43.2	195	195	195	195	195	195	195	195	195	195	195	195	195	195	195	195
D9S1777	64.7	43.2	235	235	235	235	–	–	–	–	–	–	235	235	235	235	235	235
D9S234	64.7	–	4	4	4	4	4	4	–	–	–	–	4	4	4	4	4	4
fhc-ca1	–	43.3	3	3	3	3	3	3	–	–	–	–	3	3	3	3	3	3
D9S1879	64.7	43.5	152	152	152	152	152	152	–	–	–	–	152	152	152	152	152	152
fhc-ca3	–	43.5	2	2	2	2	2	2	–	–	–	–	2	2	2	2	2	2
D9S1787	64.7	43.5	288	288	288	288	288	288	288	288	288	288	288	288	288	288	288	288
D9S1800	64.7	44.1	240	240	240	240	240	240	240	240	240	240	240	240	240	240	240	240
fhc-ca6	–	44.3	1	1	1	1			–	–	–	–	1	1	1	1	1	1
D9S273	65.8	44.6	217	217	217	217	217	217	–	–	–	–	217	217	217	217	217	217
fhc-ga2	–	44.7	9	9	9	9	9	9	–	–	–	–	9	9	9	9	9	9
D9S166	66.3	45.2	247	247	247	247	247	247	–	–	–	–	247	247	247	247	247	247
D9S1799	66.3	45.4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D9S301	66.3	45.8	231	231	231	231	–	–	–	–	–	–	231	231	235	231	235	231
fhc-ca8	–	45.9	1	1	1	1	7	1	–	–	–	–	1	7	1	1	1	1
fhc-ca9	–	46.1	7	7	7	7	7	7	–	–	–	–	7	7	7	7	7	7
fhc-ca10	–	46.1	1	1	1	1	1	1	–	–	–	–	1	1	1	1	1	1
D9S1806	66.3	46.2	264	264	264	264	260	264	–	–	–	–	264	260	264	264	264	264
fhc-ga4	–	46.2	5	5	5	5	2	5	–	–	–	–	5	2	5	5	5	5
D9S1822	66.9	46.9	159	159	159	159	–	–	–	–	–	–	159	159	159	159	159	159
D9S175	70.3	50.0	270	270	282	270	–	–	–	–	–	–	280	280	270	270	270	270
D9S1834	73.0	51.4	189	189	189	189	–	–	–	–	–	–	193	189	189	189	189	189
D9S1674	76.7	51.9	223	231	231	231	–	–	–	–	–	–	223	231	231	237	231	227

Fig. 2 Linkage disequilibrium and haplotypes in 9q12–q13. *TJP2* lies between *D9S1787* and *D9S1800*. ***a***, Results of linkage disequilibrium analysis using the 10 disease chromosomes used for the initial mapping. λ is an estimate of the excess of the associated allele on disease chromosomes relative to its frequency in the population. *P* values are given for significant associations. ***b***, Results of haplotype analysis in the 12 individuals who were genotyped. Shown are the chromosomes that share the 9q12–q13 haplotype. The shared haplotype is highlighted in gray. Missing genotype data are indicated with a minus sign (–). Also shown are genetic (Marshfield Genetics) and physical locations (Celera Discovery Systems).

are poorly anchored to tight junctions, causing greater junction permeability and depth.

Mutations in *TJP2* do not account for all FHC cases. In four individuals with FHC (9e–9g and 10c) who did not have the *TJP2* mutation, we identified an IBD haplotype in 9q22–q32, a region containing the candidate gene *BAAT*. Genomic sequencing of exons and exon-intron boundaries of *BAAT* in individual 9f identified a mutation, 226A→G, predicted to change a highly conserved methionine to valine (M76V).) Screening of all 17 individuals with FHC identified five individuals (in three families) who were homozygous with respect to this mutation ($BAAT^{226G/226G}$). Several affected individuals who were $TJP^{143C/143C}$ were also $BAAT^{226G/+}$ (Table 1). The mutation was not seen on 182 control chromosomes from Caucasian individuals, 36 control chromosomes from Amish (non-Lancaster County) individuals or 15 non-transmitted parental chromosomes. It was seen on 1 of 104 control chromosomes from Lancaster County Old Order Amish individuals.

Because BAAT mediates conjugation of bile acids with the amino acids glycine and taurine[3], we investigated serum bile acid conjugation in individuals with FHC and in control individuals (Fig. 4). The $BAAT^{226G}$ allele was associated with lower levels of conjugation in a dose-dependent fashion. $BAAT^{226G/226G}$ individuals had no amino acid–conjugated bile acids. Three individuals with elevated serum bile acid concentrations but no amino acid–conjugated bile acids have been previously reported, but no genetic analysis was done[15].

Unconjugated bile acids are not transported by the bile salt export pump, which translocates bile acids into bile[16]. In $BAAT^{226G/226G}$ individuals, who have only unconjugated bile acids, we infer that little bile acid enters bile and that most bile acid backdiffuses into plasma from the hepatocytes, leading to high serum bile acid concentrations and low intestinal concentrations.

We postulate that in $BAAT^{226G/226G}$ individuals, bile acids do not traverse hepatocytes into bile. In contrast, we believe that in

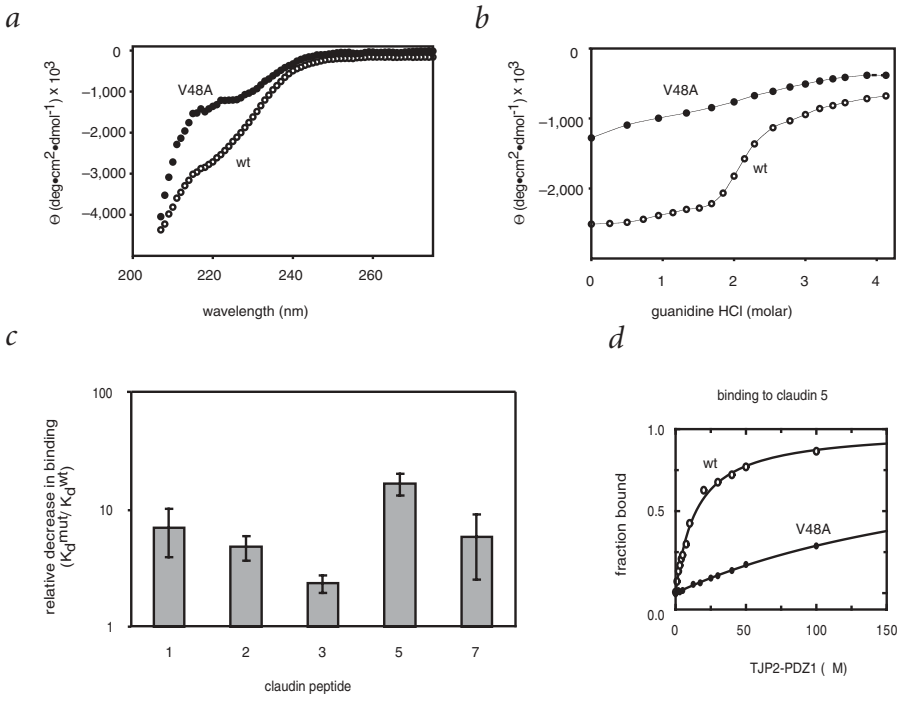

Fig. 3 Comparison of wild-type and V48A mutant TJP2-PDZ1 peptides. ***a***, Circular dichroism ellipticity versus wavelength for wild-type (wt; open circles) and V48A mutant (closed circles) TJP2-PDZ1 domains. ***b***, Chemical denaturation of wild-type (wt; open circles) and V48A mutant (closed circles) TJP2-PDZ1 domains by guanidine hydrochloride, assayed by circular dichroism spectroscopy at 222 nm. ***c***, Decrease in binding of V48A mutant TJP2-PDZ1 (relative to wild-type) to peptide C termini of claudins 1, 2, 3, 5 and 7. Binding affinity of wild-type TJP2-PDZ1 for claudin peptide C termini was as follows: claudin 1, 21 ± 3 M; claudin 2, 25 ± 6 M; claudin 3, 19 ± 2 M; claudin 5, 15 ± 2 M; and claudin 7, 35 ± 5 M. ***d***, Binding curve of wild-type (wt) and V48A mutant TJP2-PDZ1 peptides for claudin 5 peptide C terminus.

Table 2 • Hepatic tight junction depth

Individual	Depth, nm	*P* value
7c	650 ± 310 (*n* = 225)	–
Control 1	480 ± 140 (*n* = 91)	2×10^{-6}
Control 2	480 ± 230 (*n* = 98)	3×10^{-6}

Depth values represent the mean ± s.d. for the number of samples indicated in parentheses. *P* values are from Bonferroni *t*-test for controls compared to individual 7c.

TJP2$^{143C/143C}$ individuals, bile acids enter bile and then leak through tight junctions into plasma. Differences in mechanism might underlie one biochemical difference: average serum γ-glutamyltranspeptidase activity was significantly lower in *BAAT*$^{226G/226G}$ than in *TJP2*$^{143C/143C}$ individuals (9 ± 2 versus 27 ± 8; $P = 0.0009$). Bile acids elute γ-glutamyltranspeptidase from biliary-tract membranes into bile; γ-glutamyltranspeptidase thence refluxes into plasma[17]. In *BAAT*$^{226G/226G}$ individuals, low levels of bile acids in bile would explain relatively low γ-glutamyltranspeptidase concentration.

One individual with FHC (12c) did not have either mutation associated with FHC and was not homozygous in either region. This suggests the existence of a third locus associated with FHC. This individual was homozygous (possibly IBD) in several regions. One region contained the candidate gene *SLC10A1* (also called *NTCP*), which encodes a hepatic bile acid transporter. We sequenced the promoter, exons and exon-intron boundaries of *SLC10A1* in individual 12c but identified no sequence changes.

Several individuals carried mutations in both genes (Table 1), suggesting oligogenic inheritance. Independent *TJP2*$^{143C/143C}$ individuals affected with FHC were significantly more likely to be *BAAT*$^{226G/+}$ than were unaffected population controls (odds ratio = 17; 95% confidence interval = 1.3–217). Heterozygosity with respect to *BAAT*^{226G} likely increases the probability that a *TJP2*$^{143C/143C}$ individual will manifest FHC. An alternative explanation is that *TJP2* and *BAAT* are in linkage disequilibrium (they lie 42 cM apart on chromosome 9) and that *TJP2*^{143C} and *BAAT*^{226G} are associated, but a χ^2 test found no evidence for this.

Two other lines of evidence suggest oligogenic inheritance. First, all *TJP2*$^{143C/143C}$ *BAAT*$^{226G/+}$ individuals were affected with FHC (4 of 4; monozygotic twins considered one individual), whereas only 6 of 9 *TJP2*$^{143C/143C}$ *BAAT*$^{+/+}$ individuals were affected. Second, the high population frequency of *TJP2*^{143C} (7%) is unexpected for a rare disease. We have identified many fewer clinically affected *TJP2*$^{143C/143C}$ individuals than the total number predicted by Hardy–Weinberg equilibrium. Hence, inheritance of FHC seems to be oligogenic. We speculate that manifestation of FHC in *TJP2*$^{143C/143C}$ individuals requires concomitant mutation in a second locus, such as *BAAT* or the third locus postulated to be associated with FHC.

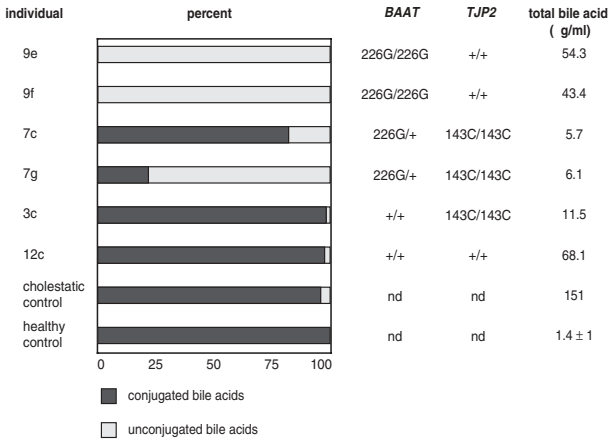

Fig. 4 Percent amino-acid conjugation of serum bile acids. Cholestatic control is an individual with genetically defined benign recurrent intrahepatic cholestasis. Healthy control represents a composite of ten healthy adult subjects: all had 100% amino-acid conjugation of serum bile acids. Subjects' genotypes at *TJP2* and *BAAT* (nd = not determined) are shown. All subjects except healthy controls were receiving UDCA.

Given the small number of *BAAT*$^{226G/226G}$ individuals, it is unclear whether genotype at *TJP2* modifies penetrance or disease expression in these individuals. The fact that one *BAAT*$^{226G/226G}$ individual is also *TJP2*$^{143C/+}$ may be due to the high population frequency of *TJP2*^{143C}.

One *TJP2*$^{143C/143C}$ *BAAT*$^{226G/+}$ individual (7c) had a nasofrontal encephalocele (a rare neural-tube defect; 1 in 35,000 live births in the West; ref. 18) and facial dysmorphism[1]. Polychaetoid (the postulated functional ortholog of TJP2 in *Drosophila melanogaster*) is involved in dorsal closure[19], a process that is probably analogous to neural-tube closure[20]. Hence, TJP2 may also have a role in neurulation.

In conclusion, FHC is associated with mutations in *TJP2* and *BAAT* that seem to affect bile acid localization. In *TJP2*$^{143C/143C}$ individuals, inheritance is probably oligogenic, with mutation in a second gene, such as *BAAT*, required for clinical disease. Our success in elucidating the genetic etiology of FHC highlights the power of population genetic approaches, in combination with the annotated human genome sequence, to identify loci and mutations associated with disease.

Methods

Affected individuals. We collected DNA and informed written consent from the individuals shown in Figure 1. For minors, we obtained consent from parents. Internal Review Boards at University of California San Francisco, Clinic for Special Children or University of Medicine and Dentistry of New Jersey approved study protocols. All individuals were of Lancaster County Old Order Amish descent. Family 7 has been previously described[1].

Absence of the *ATP8B1* 923G→T mutation in individuals with FHC. In the Amish, the ATP8B1 923G→T mutation is associated with the liver disease progressive familial intrahepatic cholestasis type 1 (PFIC1; ref. 21). We investigated the possibility that this mutation in ATP8B1 might be associated with FHC. In all 10 families with FHC that we tested (families 1–7, 9, 10 and 12), none of the affected individuals inherited the ATP8B1 923G→T mutation and none of the parents tested carried the mutation. The entire ATP8B1 coding sequence and associated exon-intron boundaries were previously sequenced in one individual with FHC (7c), and no mutations were identified (L. Klomp *et al.*, manuscript in preparation).

Genetic controls. Non-Amish Caucasian controls were from the US National Institute of General Medical Sciences Human Genetic Cell Repository Human Variation Panels (Coriell Cell Repository). Lancaster County Old Order Amish controls were individuals (or parents of individuals) with genetically diagnosed disorders. Amish (non-Lancaster County) controls were parents of individuals with PFIC1 (*ATP8B1*$^{923T/923T}$). The Lancaster County Old Order Amish are genetically distinct from other Amish populations[22].

Genome screen. We carried out a whole-genome screen on individuals indicated with an asterisk in Figure 1. Because of uncertainties in defining this newly identified disease, we used a conservative approach, limiting our initial analysis of the genome screen to data from the five affected individuals (marked with a double dagger in Figure 1) who met a narrow disease definition. These individuals each had at least two extremely high serum bile acid measurements (typically above 35 g ml^{-1}; normal is below 5 g ml^{-1}), with at least one measurement obtained at age 1 year or older, and did not have jaundice. These criteria were the consensus of three clinicians (D.H.M., G.S. and B.L.S.).

We initially used markers from the ABI PRISM Linkage Mapping Set MD-10 (approximately 400 markers) and analyzed data to identify two-marker haplotypes shared by a large proportion of chromosomes from individuals marked with a double dagger in Figure 1. Because this analysis did not generate any results of interest, we carried out a high-resolution genome screen using additional markers from the ABI PRISM Linkage Mapping Set HD-5. In total, we genotyped 753 autosomal dinucleotide markers (spaced an average of 1 every 5 cM). We also genotyped 19 X-chromosome markers and visually inspected the data but did not carry out a formal analysis, as inheritance of FHC did not seem to be X-linked. We combined data for all 753 autosomal genome screen markers and analyzed the data set to identify two-marker haplotypes on chromosomes of affected individuals. We further studied regions highlighted by this analysis by genotyping additional markers to determine if sharing was IBD. After we identified an IBD region in 9q12–q13 (which contains *TJP2* between the markers *D9S1787* and *D9S1800*), we genotyped markers from this region in additional affected individuals and family members.

We also analyzed data from roughly the first 400 markers to identify single markers of potential interest, that is, those with a highly shared allele (shared in 7 or more of 10 chromosomes from affected individuals). This analysis highlighted approximately 40 regions, which we studied further by genotyping additional markers to determine if the regions might be shared IBD.

To identify the second locus associated with FHC, we reanalyzed the genome screen data for shared two-marker haplotypes in three individuals (9e, 9f and 12c) who did not share the 9q12–q13 haplotype. We further investigated shared haplotypes in these individuals and two others (9g and 10c) for IBD sharing.

Statistics. We evaluated linkage disequilibrium between FHC and markers in 9q12–q13 using a likelihood-ratio test[23]. For this analysis, the 'disease chromosomes' were the ten chromosomes present in affected individual who were analyzed in the initial genome screen (marked with a double dagger in Figure 1) and the 'control chromosomes' were all non-transmitted parental chromosomes.

We investigated clinical and biochemical differences between FHC subtypes ($TJP2^{143C/143C}$ individuals versus $BAAT^{226G/226G}$ individuals) with a *t*-test and a Bonferroni correction for multiple testing. The only significant *P* value found was for serum γ-glutamyltranspeptidase.

We used an odds ratio to compare the number of affected $TJP2^{143C/143C}$ individuals who were also $BAAT^{226G/+}$ versus the number of unaffected Lancaster County Old Order Amish controls who were $BAAT^{226G/+}$. We found that 2 of 8 independent affected $TJP2^{143C/143C}$ individuals were also $BAAT^{226G/+}$. (The affected siblings in family 7 are not independent. Because they all share the same genotype, only one was included in the odds ratio calculation.) Only 1 of 52 control individuals was $BAAT^{226G/+}$, and this individual was $TJP2^{+/+}$.

We carried out two χ^2 tests to evaluate linkage disequilibrium between *TJP2* and *BAAT*, one using only chromosomes from Lancaster County Old Order Amish control individuals ($P = 0.79$) and one using chromosomes from these controls and affected individuals ($P = 0.25$). We identified three independent chromosomes (in families 7, 8 and 11) that carried $TJP2^{143C}$ and $BAAT^{226G}$. We noted that in family 7, the only one of these families in whom phase of the doubly heterozygous parent could be determined, the parent carried both mutations on different copies of chromosome 9.

Hardy–Weinberg equilibrium predicts approximately 136 $TJP2^{143C/143C}$ individuals in the Lancaster County Amish population. We have identified 10 affected $TJP2^{143C/143C}$ individuals (the monozygotic twins are counted as one individual). But this analysis does not take into account the fact that many affected individuals become asymptomatic by approximately age 5. Hence, we carried out a second analysis considering only children under age 5. Hardy–Weinberg equilibrium predicts approximately 25 $TJP2^{143C/143C}$ children under age 5 in the Lancaster County Amish population. We identified only 4 $TJP2^{143C/143C}$ children currently under age 5 who have ever had FHC symptoms.

There may be affected $TJP2^{143C/143C}$ individuals in the population whom we have not identified. However, the Clinic for Special Children is located in Lancaster County and specializes in treating Amish and Mennonite children who have rare metabolic disorders. Individuals with unusual disorders are frequently referred to the clinic by other pediatricians in the area.

PCR analysis. The University of California San Francisco Genomics Core Facility genotyped some markers (ABI PRISM Linkage Mapping markers and a few additional markers) with fluorescently labeled primers using standard PCR conditions and an ABI PRISM 3700 DNA analyzer. We labeled primers for all other markers radioactively and amplified markers using standard PCR conditions. We separated products by electrophoresis on denaturing polyacrylamide gels, and two independent observers scored the resulting autoradiographs. We developed several new microsatellite markers. Primer sequences are available on request.

Sequencing and mutation detection. We amplified genomic DNA of *ATP8B1* (exon 9 only), *TJP2* (all 23 coding exons), *BAAT* (all 3 coding exons) and *SLC10A1* (all 5 coding exons and promoter) using standard PCR protocols. The University of California San Francisco Genomics Core Facility sequenced PCR products using standard protocols and an ABI PRISM 3700 DNA analyzer.

The 143T→C mutation in *TJP2* and the 226A→G mutation in *BAAT* create restriction enzyme sites (*Fnu*4HI and *Nco*I, respectively). We assayed for these mutations using either sequencing (for DNA from affected individuals and Lancaster County Old Order Amish controls) or digestion of PCR products and visualization of digestion products on an agarose gel (for DNA from parents, other Amish controls and Caucasian controls). All other variants were evaluated only by sequencing.

Cloning and plasmid construction. We used PCR to clone a region of *TJP2* corresponding to the PDZ1 domain (residues 33–127 of TJP2A) from a human placental cDNA library and ligated PCR products into the expression vector pBH4 (ref. 24), which incorporates an N-terminal His_6 affinity tag. We introduced the nucleotide change corresponding to the FHC *TJP2* mutation using a standard two-step PCR method[25] and ligated PCR products into the expression vector.

Peptide synthesis. We synthesized peptides corresponding to the six C-terminal amino acids from claudins 1, 2, 3, 5 and 7 labeled with fluorescein isothiocyanate at the N terminus using conventional solid-phase fluorenylmethoxycarbonyl–amino acid chemistry on an ABI 432A Synergy Peptide Synthesizer, using Fmoc-Val-NovaSyn TGA resin (Novabiochem). Final deprotection was done on the machine. N-terminal labeling with fluorescein isothiocyanate was done by treating resin with 10 equivalents 5-(and-6)-carboxyfluorescein, succinimidyl ester (Molecular Probes) and triethylamine in dimethylformamide overnight at ambient temperature. Peptide was cleaved from resin and purified as described[25]. We verified molecular mass of all peptides by matrix-assisted laser desorption/ionization–time of flight mass spectrometry and made final stocks in filtered deionized water. Peptide sequences are as follows (from N terminus to C terminus): claudin 1, SGKDYV; claudin 2, SLTGYV; claudin 3, DRKDYV; claudin 5, DKKNYV; and claudin 7, SSKEYV.

Protein expression and purification. We transformed plasmids by heat-shock into the *Escherichia coli* strain BL21 (DE3) pLYS-S. We purified expressed proteins with Ni^{2+}-NTA resin (Qiagen) to >95% purity as described[25] and dialyzed samples extensively against 50 mM sodium phosphate, 100 mM NaCl, pH 7.5 buffer. The expressed TJP2-PDZ1 domain construct was not amenable to freeze-thaw and was therefore used fresh (immediately after dialysis) for biochemical assays.

Circular dichroism spectroscopy. We carried out circular dichroism spectroscopy using a JASCO 715 spectropolarimeter. Samples of wild-type or V48A TJP2-PDZ1 (70 M) were in 0.1-cm quartz cuvettes at 10 °C in 50 mM sodium phosphate, 100 mM NaCl, pH 7.5 buffer. We collected data at 1-nm intervals with 1 s averaging time per data point and a total of 5 repeats per scan.

We did chemical denaturation studies at 20 °C in 1.0-cm–path length quartz cuvettes with wild-type or V48A TJP2-PDZ1 (35 M) in 50 mM sodium phosphate, 100 mM NaCl, pH 7.5 buffer. We manually added a stock solution of 8 M guanidine hydrochloride at each data point and collected data at 222 nm for 100 s.

Binding studies. We measured binding affinities of peptides N-terminally labeled with fluorescein isothiocyanate to wild-type and mutant TJP2-

PDZ1 domains by fluorescence polarization with a LJL Biosystems Analyst AD fluorescence multiwell plate reader and analyzed data with LJL Biosystems Criterion Host software. Each data point was 100 l total volume with a peptide concentration of 5 nM. Data were fit to a bimolecular binding equation as described[25] using the program ProFit 5.1.0 (Quantum Soft). All binding experiments were done in 50 mM sodium phosphate, pH 7.5, 100 mM NaCl buffer at 10 °C. Each K_d value is an average derived from four independent experiments.

Ultrastructural studies. We placed liver tissue from individual 7c (obtained during laparotomy conducted for diagnostic purposes) in phosphate-buffered glutaraldehyde fixative solution, postfixed it, stained it *en bloc* with osmium tetroxide and processed it by standard techniques into resin. We stained ultrathin sections with uranyl acetate–lead citrate and examined sections by transmission electron microscopy. We exposed 50 micrographs of canaliculi and measured tight junction depth. No material appropriate for freeze-fracture analysis was available; hence, we were unable to count the number of strands present in junctions.

We also measured junction depth in 50 micrographs each of similarly processed tissue from two control liver biopsy specimens. Because healthy liver tissue was unavailable (donor organs, often used as controls in liver studies, are frankly abnormal on light microscopy and are unsuitable for ultrastructural study), we used tissue from two children with mild liver disease characterized by slight clinical-biochemical evidence of hepatocellular injury ('transaminitis') without cholestasis.

Serum bile acid analysis. We combined 2 ml of serum with 2 ml distilled water and an internal standard (20 g nor-deoxycholic acid in 100 l methanol) and measured bile acids with gas liquid chromatography (GLC) as previously described[1] with the following modifications. After washing, we divided the sample into two aliquots and did enzymatic hydrolysis on only one sample. We used the hydrolyzed sample to determine total bile acids and the non-hydrolyzed sample to determine levels of unconjugated bile acids. The error of percent conjugation (determined by running a standard sample five times) was less than 5%. We used a maximum oven temperature of 262 °C.

The retention times for various bile acids relative to nor-deoxycholic acid as methyl ester–trimethylsilyl ether derivatives (retention time, 14.14 m) were as follows: lithocholic acid, 1.09; deoxycholic acid, 1.20; chenodeoxycholic acid, 1.26; cholic acid, 1.28; ursodeoxycholic acid, 1.34; ursocholic acid, 1.40. We confirmed bile acid structure by mass-spectral fragmentation patterns generated with a Hewlett Packard 5972A mass selective detector coupled to a 6890 gas chromatograph using a CP-Sil 5 CB GLC column.

Healthy controls were all non-Amish adults with no symptoms of liver disease. The cholestatic control was a non-Amish patient with benign recurrent intrahepatic cholestasis; the sample was taken during a cholestatic episode.

URL. Marshfield Genetics, http://www.marshfieldclinic.org/research/genetics.

Acknowledgments

The authors thank the individuals with FHC and their family members for their participation; I. Herskowitz for support and contributions; S. Service for software and advice; C. Hendrickson, J. Doherty, J. Woo, H. Yoon, E. Song, J. Vargas, R. Senkus, M. Maksimak, E. Cunningham, D. Agard, A. Frankel, V. Calabro, K. Novak and R. K. Guy for advice, assistance and the use of equipment; and N. Risch, N. Freimer, C. Slates, W. Hsueh, A. Slavotinek, R. Kelley and F. Collins for discussions. This work was supported by grants from the US National Institutes of Health, the Howard Hughes Medical Institute, the University of California San Francisco Program in Human Genetics and the Tetrad Graduate Program.

Competing interests statement

The authors declare that they have no competing financial interests.

Received 19 December 2002; accepted 28 March 2003.

1. Morton, D.H. *et al*. Abnormal hepatic sinusoidal bile acid transport in an Amish kindred is not linked to *FIC1* and is improved by ursodiol. *Gastroenterology* **119**, 188–195 (2000).
2. Shneider, B.L. *et al*. Hepatic basolateral sodium-dependent-bile acid transporter expression in two unusual cases of hypercholanemia and in extrahepatic biliary atresia. *Hepatology* **25**, 1176–1183 (1997).
3. Falany, C.N., Johnson, M.R., Barnes, S. & Diasio, R.B. Glycine and taurine conjugation of bile acids by a single enzyme. Molecular cloning and expression of human liver bile acid CoA:amino acid N-acyltransferase. *J. Biol. Chem.* **269**, 19375–19379 (1994).
4. Letunic, I. *et al*. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.* **30**, 242–244 (2002).
5. Harris, B.Z. & Lim, W.A. Mechanism and role of PDZ domains in signaling complex assembly. *J. Cell Sci.* **114**, 3219–3231 (2001).
6. Itoh, M. *et al*. Direct binding of three tight junction-associated MAGUKs, ZO-1, TJP2, and ZO-3, with the COOH termini of claudins. *J. Cell Biol.* **147**, 1351–1363 (1999).
7. Mitic, L.L., Van Itallie, C.M. & Anderson, J.M. Molecular physiology and pathophysiology of tight junctions I. Tight junction structure and function: lessons from mutant animals and proteins. *Am. J. Physiol. Gastrointest. Liver Physiol.* **279**, G250–G254 (2000).
8. Jesaitis, L.A. & Goodenough, D.A. Molecular characterization and tissue distribution of ZO-2, a tight junction protein homologous to ZO-1 and the *Drosophila* discs-large tumor suppressor protein. *J. Cell Biol.* **124**, 949–961 (1994).
9. Tsukita, S., Furuse, M. & Itoh, M. Multifunctional strands in tight junctions. *Nat. Rev. Mol. Cell Biol.* **2**, 285–293 (2001).
10. McCarthy, K.M. *et al*. Inducible expression of claudin-1-myc but not occludin-VSV-G results in aberrant tight junction strand formation in MDCK cells. *J. Cell Sci.* **113** Pt 19, 3387–3398 (2000).
11. Heiskala, M., Peterson, P.A. & Yang, Y. The roles of claudin superfamily proteins in paracellular transport. *Traffic* **2**, 93–98 (2001).
12. Dawson, P.A. Bile Secretion and the Enterohepatic Circulation of Bile Acids. in *Gastrointestinal and Liver Disease* (eds. Feldman, M., Friedman, L.S. & Sleisenger, M.H.) 1051–1062 (Saunders, Philadelphia, 2002).
13. Rahner, C., Stieger, B. & Landmann, L. Structure–function correlation of tight junctional impairment after intrahepatic and extrahepatic cholestasis in rat liver. *Gastroenterology* **110**, 1564–1578 (1996).
14. Koga, A. & Todo, S. Morphological and functional changes in the tight junctions of the bile canaliculi induced by bile duct ligation. *Cell Tissue Res.* **195**, 267–276 (1978).
15. Setchell, K.D. & O'Connell, N.C. Disorders of Bile Acid Synthesis and Metabolism: A Metabolic Basis for Liver Disease. in *Liver Disease in Children* (eds. Suchy, F.J., Sokol, R.J. & Balistreri, W.F.) 701–734 (Lippincott Williams & Wilkins, Philadelphia, 2001).
16. Noe, J., Stieger, B. & Meier, P.J. Functional expression of the canalicular bile salt export pump of human liver. *Gastroenterology* **123**, 1659–1666 (2002).
17. Ronchi, G. & Desmet, V.J. Histochemical study of gamma glutamyl transpeptidase (GGT) in experimental intrahepatic and extrahepatic cholestasis. *Beitr. Pathol.* **150**, 316–321 (1973).
18. Mahapatra, A.K. & Suri, A. Anterior encephaloceles: a study of 92 cases. *Pediatr. Neurosurg.* **36**, 113–118 (2002).
19. Chen, C.M. *et al*. Polychaetoid is required to restrict segregation of sensory organ precursors from proneural clusters in *Drosophila*. *Mech. Dev.* **57**, 215–227 (1996).
20. Colas, J.F. & Schoenwolf, G.C. Towards a cellular and molecular understanding of neurulation. *Dev. Dyn.* **221**, 117–145 (2001).
21. Bull, L.N. *et al*. A gene encoding a P-type ATPase mutated in two forms of hereditary cholestasis. *Nat. Genet.* **18**, 219–224 (1998).
22. Hostetler, J.A. *Amish Society* (The Johns Hopkins University Press, Baltimore, 1993).
23. Terwilliger, J.D. A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *Am J. Hum. Genet.* **56**, 777–787 (1995).
24. Hillier, B.J., Christopherson, K.S., Prehoda, K.E., Bredt, D.S. & Lim, W.A. Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. *Science* **284**, 812–815 (1999).
25. Harris, B.Z., Hillier, B.J. & Lim, W.A. Energetic determinants of internal motif recognition by PDZ domains. *Biochemistry* **40**, 5921–5930 (2001).