

Report

A Missense Mutation (R565W) in *Cirhin* (FLJ14728) in North American Indian Childhood Cirrhosis

Pierre Chagnon,¹ Jacques Michaud,¹ Grant Mitchell,¹ Jocelyne Mercier,¹ Jean-François Marion,¹ Eric Drouin,² Andrée Rasquin-Weber,² Thomas J. Hudson,³ and Andrea Richter¹

Divisions of ¹Medical Genetics and ²Gastroenterology, Hôpital Ste-Justine, Centre Hospitalier Universitaire Mère-Enfant, Université de Montréal, and ³Montréal Genome Centre, McGill University Health Centre Research Institute, Montréal

North American Indian childhood cirrhosis (CIRH1A, or NAIC), a severe autosomal recessive intrahepatic cholestasis described in Ojibway-Cree children from northwestern Quebec, is one of several familial cholestases with unknown molecular etiology. It typically presents with transient neonatal jaundice, in a child who is otherwise healthy, and progresses to biliary cirrhosis and portal hypertension. Clinical and physiological investigations have not revealed the underlying cause of the disease. Currently, liver transplantation is the only effective therapy for patients with advanced disease. We previously identified the NAIC locus by homozygosity mapping to chromosome 16q22. Here we report that an exon 15 mutation in gene FLJ14728 (alias *Cirhin*) causes NAIC: c.1741C→T in GenBank cDNA sequence NM_032830, found in all NAIC chromosomes, changes the conserved arginine 565 codon to a tryptophan, altering the predicted secondary structure of the protein. *Cirhin* is preferentially expressed in embryonic liver, is predicted to localize to mitochondria, and contains WD repeats, which are structural motifs frequently associated with molecular scaffolds.

North American Indian childhood cirrhosis (CIRH1A or NAIC [MIM 604901]) is a distinct form of severe familial cholestasis of unknown etiology (Arnell et al. 1997; Drouin et al. 2000a; Jacquemin 2000; Chen et al. 2001) described in Ojibway-Cree aboriginal children in the Abitibi region of northwestern Quebec (Drouin et al. 2000b). The disease typically presents with transient neonatal jaundice, in a child who is otherwise well, and progresses to biliary cirrhosis and portal hypertension. Biochemical and histopathological features of the disease suggest involvement of the bile ducts rather than of the bile canaliculi. They include elevated gamma glutamyl transferase and alkaline phosphatase levels, and, typically, marked fibrosis around portal bile ducts. Clinical and physiological investigations have not revealed the underlying cause of NAIC. Currently, liver transplan-

tation is the only effective therapy for patients with advanced disease.

In a previous study (Betard et al. 2000), we localized the gene responsible for autosomal recessive intrahepatic cholestasis NAIC/CIRH1A to chromosome 16q22, through use of a genomewide scan. A common 4.9-cM haplotype was found to be shared and homozygous in all patients with NAIC and was flanked on the centromeric side by D16S3398 and on the telomeric side by D16S512 (Betard et al. 2000). Analysis of additional markers in an enlarged cohort allowed us to narrow this conserved ancestral haplotype to a 2.4-Mb region between D16S3085 and D16S3095 (fig. A [online only]). According to the Human Genome Project Working Draft (UCSC Genome Bioinformatics Web site), this region contains 45 genes of known and unknown function.

Using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems), we sequenced the coding region, splice junctions, flanking intronic sequence, and ~1 kb of the putative promoters of all 27 genes known to be expressed in liver (on the basis of EST sources listed in UniGene) (fig. A [online only]). Nineteen of these were located between the two recombinant markers D16S3085 and D16S3095. Initially, six samples were sequenced (patient DNA pool, obligate heterozygote DNA pool, two

Received July 25, 2002; accepted for publication September 3, 2002; electronically published November 4, 2002.

Address for correspondence and reprints: Dr. Andrea Richter, Service de Génétique Médicale, Hôpital Sainte-Justine, 3175 Côte Sainte-Catherine, Montréal, Québec, Canada, H3T 1C5. E-mail: andrea.richter@umontreal.ca

© 2002 by The American Society of Human Genetics. All rights reserved. 0002-9297/2002/7106-0020\$15.00

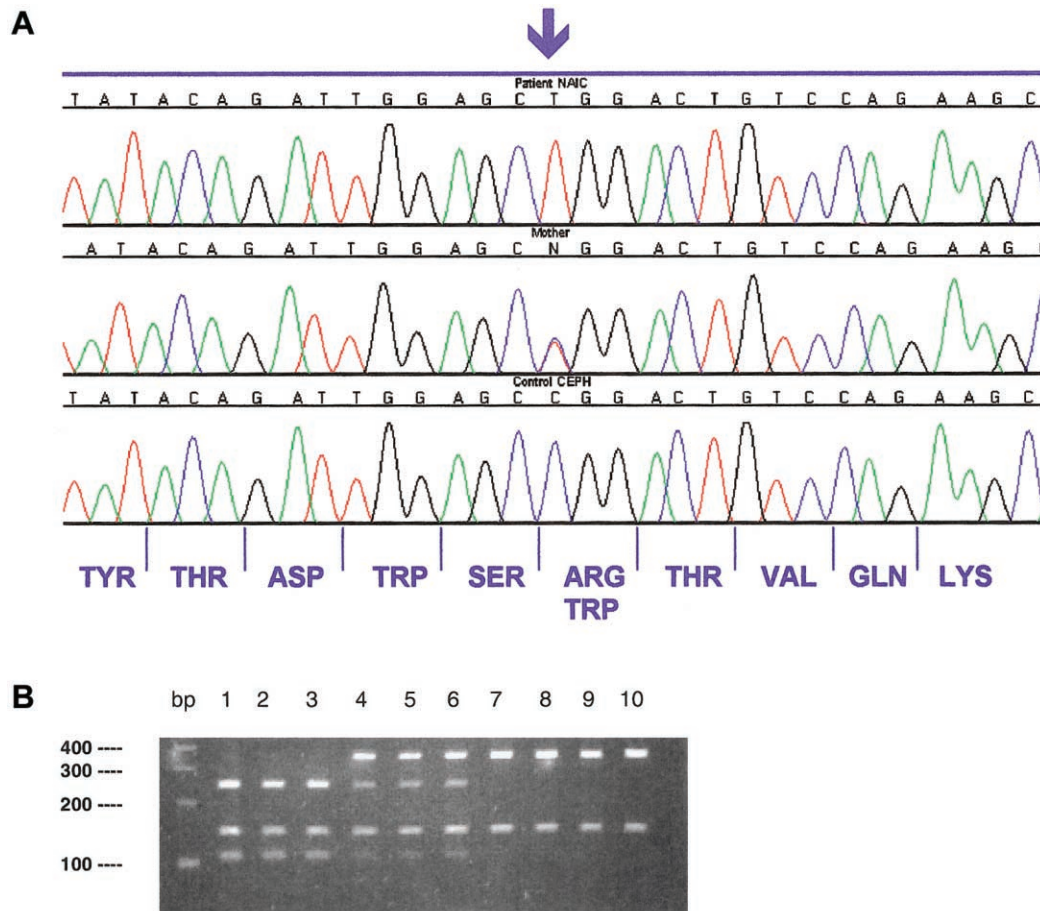


Figure 1 Mutation detection. *A*, Electropherogram of patient (*top*), parent (*middle*), and control subject (*bottom*). At the bottom, the amino acid sequence corresponding to this DNA region is shown. The NAIC mutation changes an arginine codon to a tryptophan codon. *B*, Detection of the NAIC mutation by a PCR-RFLP test. The NAIC mutation creates an *AluI* restriction site. Exon 15 of the FLJ14728 gene was amplified by PCR using forward primer 5'-AAT GCC CTT CTG GTC TTT CTG and reverse primer 5'-CGC AAA GTA GAC AAA TAG TCA. PCR was performed under the following conditions: 95°C for 5 min, followed by 35 cycles of 95°C for 45 s, 58°C for 30s, and 72°C for 45s. The amplicon was digested with *AluI*, and the digestion products were separated on 3.0% agarose gels. The PCR fragment also contains a constant internal *AluI* site that serves as a positive control for digestion and that yields a 150-bp fragment. *Lanes 1–3*, patients with NAIC; *lanes 4–6*, heterozygous parents; *lanes 7–10*, homozygous normal individuals.

NAIC individual patient DNAs, a DNA sample from a patient's mother, and a CEPH control [individual NA07057, CEPH/Utah pedigree 1331 from CORIELL]). Informative variants (heterozygous parents and homozygous patients) were genotyped by sequencing or by PCR-RFLP in all samples from the five families with NAIC (22 patients, 22 parents, and 30 siblings).

Multiple alignments of DNA sequences, obtained using Sequencher software (version 4.0.5; Gene Codes), allowed us to identify 88 SNPs, 57 of which were located between D16S3085 and D16S3095. Eighty-seven SNPs either were located in introns or were neutral exonic variants.

Only one gene inside the NAIC critical region (encoding hypothetical protein FLJ14728; UniGene cluster

Hs.151001) contained a missense mutation that could have a functional effect and could be the cause of NAIC. This mutation was located in exon 15 of the cDNA sequence NM_032830 (c.1741C→T) or in BAC genomic sequence AC009131 (g.59305C→T) (fig. 1A) and substitutes a tryptophan for a conserved arginine at codon 565 (R565W). FLJ14728 has 17 exons spanning 36.3 kb of genomic sequence and is flanked by the syntrophin, beta 2 (*SNTB2*), and hyaluronan synthase 3 (*HAS3*) genes (fig. A [online only]). The transcript is 2,192 bp with a single ORF of 2,060 bp encoding a 686-amino acid protein (GenBank accession number NP_116219). We named the gene encoding FLJ14728 "Cirhin."

We confirmed the association of this mutation with the transmission of NAIC by sequencing 74 individuals



Figure 2 Partial amino acid sequences of Cirhin and its homologues in vertebrates. The underlined arginine (R) is the residue mutated in patients with NAIC. Multiple sequence alignment was performed using the ClustalW alignment tool with Blosum weight matrix at the Baylor College of Medicine Search Launcher Web site. To generate the output, we used BoxShade 3.21 (BoxShade Server Web site). All homologous protein sequences (GenBank ID numbers shown) were originally identified by tblastn analysis at the NCBI BLAST Home Page, using the FLJ14728 protein sequence (GenBank accession number NP_116219) as a query against EST databases. In the residue numbering, 1 is the first amino acid in each of the predicted proteins.

from the five families with NAIC. All patients ($n = 22$) were T/T homozygotes. The observation that the homozygous missense mutation was absent in all unaffected family members, since all parents ($n = 22$) were heterozygous (T/C), and that the 30 unaffected siblings were either heterozygous (21/30) or C/C homozygous (9/30), lends support to our assertion that this is a disease-causing mutation.

To determine the frequency of the NAIC c.1741C→T mutation in different populations, we genotyped 367 control DNAs by means of a simple PCR-RFLP assay, which exploits the presence of an *AluI* restriction site created by the mutant allele (fig. 1B). This PCR-RFLP assay will be particularly useful as a mutation screening test in at-risk families. The DNA samples were obtained from 122 unrelated individuals from the province of Quebec, 120 CEPH parents, 56 Ojibway-Cree Indians with no known relationship to the families with NAIC, 25 Nadene Indians, 10 Mexican Americans, 10 South American Indians, and 24 DNAs from the M24PDR panel obtained from Coriell Cell Repositories (with DNA samples from European Americans, African Americans, Mexican Americans, Asian Americans, and Native Americans).

Only a single c.1741C→T allele was discovered in a Mexican American individual. Investigation of SNPs in genes closely linked to *Cirhin*—including the nuclear factor of activated T cells, cytoplasmic, calcineurin-dependent 3 (*NFATC3*), cadherin 3 (*CDH3*), *SNTB2*, and vacuolar protein sorting 4A (*VPS4*) genes—indicates that the c.1741C→T variation observed in this individual is on a different haplotype (data not shown), pointing to an origin that is likely to be independent from that of the Ojib-

way-Cree NAIC mutation. Thus, our identification of 1 mutant allele among 734, including at least 182 alleles from Native Americans, suggests that the mutation arose spontaneously in this sample and was not the result of Native American admixture. Of note, the normal C nucleotide is within a CpG dinucleotide, a known mutational hotspot (Templeton et al. 2000). In spite of this, the demonstration of a R565W *Cirhin* mutation in a non-Ojibway-Cree individual suggests that the *Cirhin* gene should be investigated in cases of nonsyndromic cholestasis clinically similar to NAIC.

In the BLAST analysis of the *Cirhin* gene sequence against all available databases, the mutant allele was absent from 101 different mammalian sequences (BAC AC009131, 64 human ESTs, 14 human cDNAs, 17 mouse ESTs, and 5 rat ESTs). This confirmed that the NAIC mutation is rare in the general population.

Cirhin has no known human paralogs. The murine ortholog is UniGene cluster Mm.10665. Using Sequencher, we assembled all mouse ESTs of this cluster, to determine the mouse coding sequence, which is 89% identical to the human sequence. In the deduced amino acid sequence of Cirhin in mice and other vertebrates, R565 is highly conserved (fig. 2).

In *Saccharomyces cerevisiae*, an ORF on chromosome IV (YDR324c) encodes a 776-residue protein of unknown function, with 23% amino acid identity to Cirhin. The amino acid position equivalent to R565 in human Cirhin is occupied by another dibasic amino acid, lysine. This suggests that the basic residue arginine may be functionally significant and that mutation to the aromatic amino acid tryptophan probably affects the function of the protein. Strains of *S. cerevisiae* deleted for

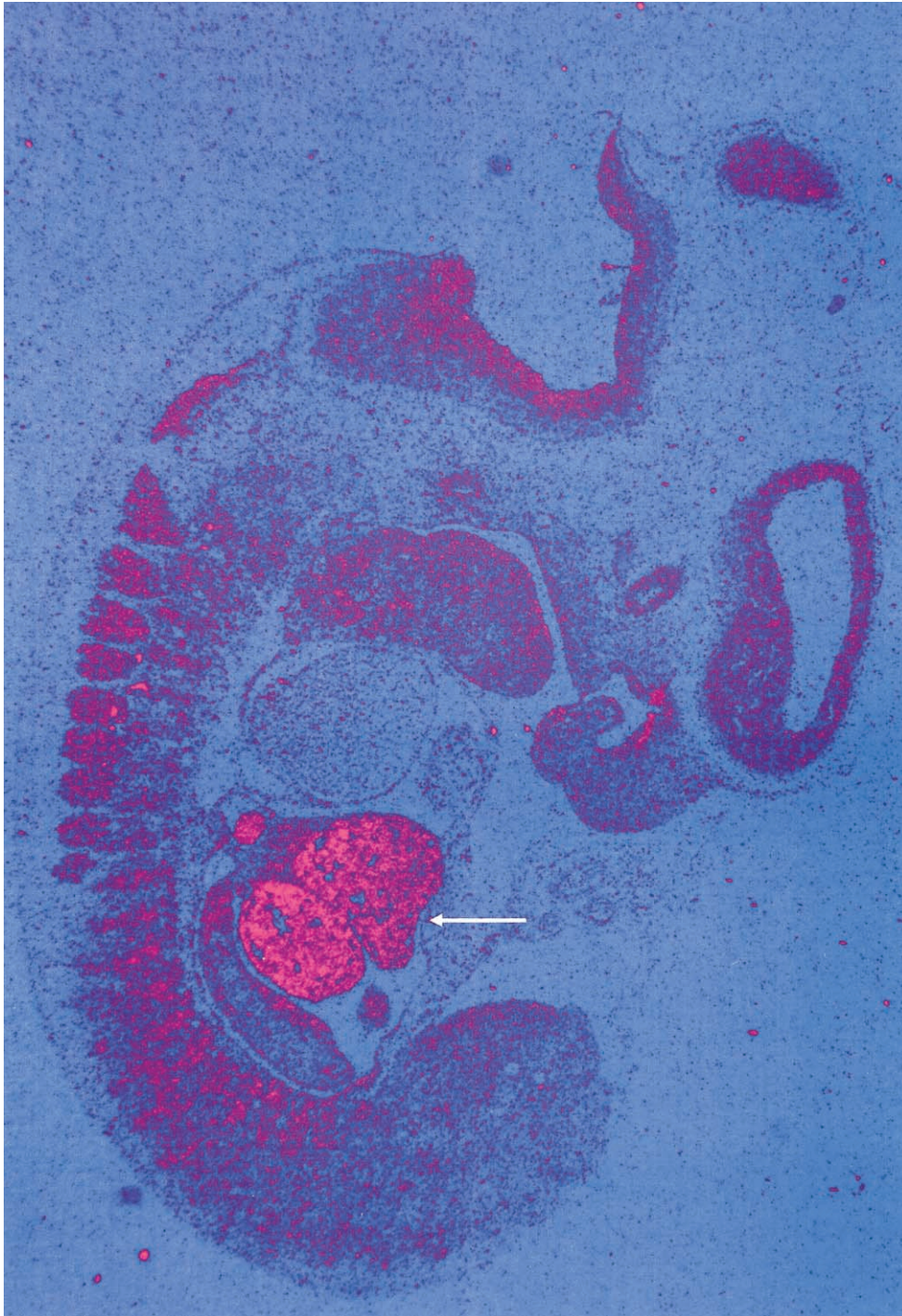


Figure 3 Embryonic expression of *Cirhin*. In situ hybridization on a sagittal section of a day-11.5 mouse embryo with a 3' probe showed prominent expression of *Cirhin* in the liver (arrow) and much weaker levels in the somites, brain, and craniofacial structures. Probe 1 specific to the 5' portion of *Cirhin* gave identical results, whereas the corresponding sense probes showed no detectable signals. The mouse primer sequences for amplification by RT-PCR of probe 1 (5' fragment of the gene) are 5'-CAG ATT GGC TGT TTC ACG AAC and TGT TGC CTA TCT AAA ACC ATC, and those for probe 2 (3' fragment) are 5'-CTT TCC AGC CGC AGT CAG GTA and 5'-CCG CCT CAA GAA GAC ATC TGA T. Total adult mouse liver RNA was used as a template for the RT-PCRs.

YDR324c are nonviable (Winzeler et al. 1999). Because of the nonviable phenotype in yeast lacking the *Cirhin* ortholog, it is possible that complete *Cirhin* deficiency may be lethal in mammals and that mutant *Cirhin* containing R565W has partial activity.

To determine the effect of the c.1741C→T allele on gene expression, we compared the intensity of a 600-nt RT-PCR fragment from the 3' portion of *Cirhin* through use of liver RNA of a patient with NAIC and a control individual. The expression levels were equivalent (data not shown). Patients with NAIC have neonatal jaundice (Drouin et al. 2000b), consistent with an abnormality of prenatal liver development. To determine whether the NAIC gene could play an important role in liver embryogenesis, we used in situ hybridization to investigate the gene expression pattern during mouse embryonic development. The two murine NAIC probes used in these experiments were generated from the 5' and 3' portions of the gene. Since there is no complete mouse homologue in GenBank, we obtained the virtual mouse sequence by comparing the human FLJ14728 sequence (GenBank accession number NM_032830) to the mouse EST database, through use of BLAST (NCBI BLAST Home Page). Two different mouse ESTs, both from UniGene cluster Mm.10665, were used as templates for designing the RT-PCR primers. Indeed, in situ hybridization in day-11.5 mouse embryos shows that *Cirhin* is highly and predominantly expressed in liver, with lower levels of expression in somites, brain, and craniofacial structures (fig. 3), whereas the corresponding sense probe showed only background signal (data not shown). *Cirhin* expression showed a significant decrease in day-12.5 embryos and in the newborn, but we observed detectable levels of mRNA in newborn and adult mouse liver by RT-PCR (data not shown).

The secondary structure of the *Cirhin* protein was predicted through use of Internet tools, including ExPASy (ExPASy Molecular Biology Server). *Cirhin* has a predicted molecular weight of 77 kD, with a theoretical isoelectric point of 9.03. iPSORT (iPSORT Home Page) predicts the presence of an N-terminal mitochondrial targeting peptide for *Cirhin*. The R565W substitution resulting from the NAIC mutation replaces an isolated α -helical domain with an extended strand conformation in the predicted 2-D structure of the protein determined at the Web-based Protein Sequence Analysis (PSA) Server and at the Network Protein Sequence Analysis at IBCP Web site. This conformation change could have functional consequences.

Cirhin is classified in NCBI LocusLink as a WD repeat-containing protein. The secondary-structure predictions used detected 10 WD repeats. WD proteins are involved in signal transduction, chromatin assembly, mRNA synthesis, RNA splicing, vascular trafficking, cytoskeletal assembly, transcription initiation complex as-

sembly, and, often, the organization of proteins into complexes (Smith et al. 1999). The R565W mutation lies outside of the predicted WD domains.

Most genes previously described as causing familial cholestases are implicated in membrane transport of bile components (Li et al. 1997; Bull et al. 1998; de Vree et al. 1998; Strautnieks et al. 1998; Jansen and Muller 2000). Because *Cirhin* has no apparent connection with biliary function, it may provide new insights into the mechanisms of cholestasis.

Molecular studies of Canadian First Nations are providing useful tools for diagnosis and therapy (Robinson et al. 1984; Greenberg et al. 1995; Weiler et al. 1996; Camacho et al. 1999; Meira et al. 2000; Hegele 2001; Mok et al. 2002; Triggs-Raine et al. 2002). To our knowledge, this is the first time that gene function has been established exclusively by the study of DNA samples of North American aboriginal populations. Conversely, the discovery that a single major mutation causes NAIC will provide a useful genetic counseling tool for individuals from at-risk communities. In addition, there are many cases of nonsyndromic cholestasis with unknown molecular defect in which the *Cirhin* gene can now be investigated.

Acknowledgments

We thank the patients and their families, for participating in the study, as well as Ms. Virginie Poisson and Ms. Julie Fortin, for technical assistance. This study was supported by Canadian Institutes of Health Research (CIHR) operating grant MOP-38050 (to A.R. and G.M.). G.M. holds a Fonds de la Recherche en Santé du Québec Chercheur-National award, and Ja.M. and T.J.H. are recipients of CIHR Clinician-Scientist Awards.

Electronic-Database Information

Accession numbers and URLs for data presented herein are as follows:

Baylor College of Medicine Search Launcher: Multiple Sequence Alignments, <http://searchlauncher.bcm.tmc.edu/multi-align/>

BoxShade Server, http://www.ch.embnet.org/software/BOX_form.html

ExPASy Molecular Biology Server, <http://ca.expasy.org/>

GenBank, <http://www.ncbi.nlm.nih.gov/Genbank> (for hypothetical protein FLJ14728 [accession number NM_032830], CA7 [accession number NM_005182], DNCL12 [accession number NM_006141], CES2 [accession number NM_003869], TRADD [accession number NM_003789], SLC9A5 [accession number NM_004594], HSD11B2 [accession number NM_000196], ATP6DV [accession number X71490], CTCF [NM_006565], PSMB10 [accession number NM_002801], LCAT [accession number NM_000229], NFATC3 [accession number NM_004555],

SMPD3 [accession number NM_018667], *CDH3* [accession number NM_001793], *CDH1* [accession number NM_004360], hypothetical protein FLJ12688 [accession number AK022750], *HAS3* [accession number NM_005329], *SNTB2* [accession number NM_006750], *VPS4* [accession number NM_013245], *LOC64146* [accession number NM_022341], hypothetical protein HSPC031 [accession number NM_016101], *TERF2* [accession number NM_005652], *CYB5-M* [accession number NM_030579], *DIA4* [accession number NM_000903], *TAT* [accession number NM_000353], *DHODH* [accession number M94065], *PRP16* [accession number NM_014003], *Homo sapiens* chromosome 16 clone RP11-502K10 [accession number AC009131], and YDR324Cp [accession number AAB64760; GI:915001])
 iPSORT Home Page, <http://www.hypothesiscreator.net/iPSORT/>
 LocusLink, <http://www.ncbi.nlm.nih.gov/LocusLink/>
 NCBI BLAST Home Page, <http://www.ncbi.nlm.nih.gov/BLAST/>
 Network Protein Sequence Analysis at IBCP, http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html
 Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>
 Protein Sequence Analysis (PSA) Structure Prediction Server, <http://bmerc-www.bu.edu/psa/>
 UCSC Genome Bioinformatics, <http://genome.ucsc.edu/>
 UniGene, <http://www.ncbi.nlm.nih.gov/UniGene/> (for hypothetical protein FLJ14728 [Hs.151001] and the murine ortholog of Cirhin [Mm.10665])

References

- Arnell H, Nemeth A, Anneren G, Dahl N (1997) Progressive familial intrahepatic cholestasis (PFIC): evidence for genetic heterogeneity by exclusion of linkage to chromosome 18q21-q22. *Hum Genet* 100:378–381
- Betard C, Rasquin-Weber A, Brewer C, Drouin E, Clark S, Verner A, Darmond-Zwaig C, Fortin J, Mercier J, Chagnon P, Fujiwara TM, Morgan K, Richter A, Hudson TJ, Mitchell GA (2000) Localization of a recessive gene for North American Indian childhood cirrhosis to chromosome region 16q22—and identification of a shared haplotype. *Am J Hum Genet* 67:222–228
- Bull LN, van Eijk MJ, Pawlikowska L, DeYoung JA, Juijn JA, Liao M, Klomp LW, Lomri N, Berger R, Scharschmidt BF, Knisely AS, Houwen RH, Freimer NB (1998) A gene encoding a P-type ATPase mutated in two forms of hereditary cholestasis. *Nat Genet* 18:219–224
- Camacho JA, Obie C, Biery B, Goodman BK, Hu CA, Al-mashanu S, Steel G, Casey R, Lambert M, Mitchell GA, Valle D (1999) Hyperornithinaemia-hyperammonaemia-homocitrullinuria syndrome is caused by mutations in a gene encoding a mitochondrial ornithine transporter. *Nat Genet* 22:151–158
- Chen HL, Chang PS, Hsu HC, Lee JH, Ni YH, Hsu HY, Jeng YM, Chang MH (2001) Progressive familial intrahepatic cholestasis with high γ -glutamyltranspeptidase levels in Taiwanese infants: role of *MDR3* gene defect? *Pediatr Res* 50:50–55
- de Vree JM, Jacquemin E, Sturm E, Cresteil D, Bosma PJ, Aten J, Deleuze JF, Desrochers M, Burdelski M, Bernard O, Oude Elferink RP, Hadchouel M (1998) Mutations in the *MDR3* gene cause progressive familial intrahepatic cholestasis. *Proc Natl Acad Sci USA* 95:282–287
- Drouin E, Mitchell GA, Rasquin-Weber A (2000a) Other inherited cholestatic disorders. In: Walker AW, Durie PR, Hamilton JR, Walker-Smith JA, Watkins JB (eds) *Pediatric gastrointestinal disease: pathophysiology, diagnosis, management*. 3rd ed. BC Decker, Hamilton, Ontario, pp 1211–1218
- Drouin E, Russo P, Tuchweber B, Mitchell G, Rasquin-Weber A (2000b) North American Indian cirrhosis in children: a review of 30 cases. *J Pediatr Gastroenterol Nutr* 31:395–404
- Greenberg CR, Reimer D, Singal R, Triggs-Raine B, Chudley AE, Dilling LA, Philipps S, Haworth JC, Seargeant LE, Goodman SI (1995) A G-to-T transversion at the +5 position of intron 1 in the glutaryl CoA dehydrogenase gene is associated with the Island Lake variant of glutaric acidemia type I. *Hum Mol Genet* 4:493–495
- Hegele RA (2001) Genes, environment and diabetes in Canadian aboriginal communities. *Adv Exp Med Biol* 498:11–20
- Jacquemin E (2000) Progressive familial intrahepatic cholestasis. Genetic basis and treatment. *Clin Liver Dis* 4:753–763
- Jansen PL, Muller M (2000) Genetic cholestasis: lessons from the molecular physiology of bile formation. *Can J Gastroenterol* 14:233–238
- Li L, Krantz ID, Deng Y, Genin A, Banta AB, Collins CC, Qi M, Trask BJ, Kuo WL, Cochran J, Costa T, Pierpont ME, Rand EB, Piccoli DA, Hood L, Spinner NB (1997) Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. *Nat Genet* 16:243–251
- Meira LB, Graham JM Jr, Greenberg CR, Busch DB, Doughty AT, Ziffer DW, Coleman DM, Savre-Train I, Friedberg EC (2000) Manitoba aboriginal kindred with original cerebro-oculo-facio-skeletal syndrome has a mutation in the Cockayne syndrome group B (CSB) gene. *Am J Hum Genet* 66:1221–1228
- Mok A, Cao H, Zinman B, Hanley AJ, Harris SB, Kennedy BP, Hegele RA (2002) A single nucleotide polymorphism in protein tyrosine phosphatase PTP-1B is associated with protection from diabetes or impaired glucose tolerance in Ojibwe. *J Clin Endocrinol Metab* 87:724–727
- Robinson BH, Oei J, Sherwood WG, Applegarth D, Wong L, Haworth J, Goodyer P, Casey R, Zaleski LA (1984) The molecular basis for the two different clinical presentations of classical pyruvate carboxylase deficiency. *Am J Hum Genet* 36:283–294
- Smith TF, Gaitatzes C, Saxena K, Neer EJ (1999) The WD repeat: a common architecture for diverse functions. *Trends Biochem Sci* 24:181–185
- Strautnieks SS, Bull LN, Knisely AS, Kocoshis SA, Dahl N, Arnell H, Sokal E, Dahan K, Childs S, Ling V, Tanner MS, Kagalwalla AF, Nemeth A, Pawlowska J, Baker A, Mieli-Vergani G, Freimer NB, Gardiner RM, Thompson RJ (1998)

- A gene encoding a liver-specific ABC transporter is mutated in progressive familial intrahepatic cholestasis. *Nat Genet* 20:233–238
- Templeton AR, Clark AG, Weiss KM, Nickerson DA, Boerwinkle E, Sing CF (2000) Recombinational and mutational hotspots within the human lipoprotein lipase gene. *Am J Hum Genet* 66:69–83
- Triggs-Raine BL, Kirkpatrick RD, Kelly SL, Norquay LD, Cattini PA, Yamagata K, Hanley AJ, Zinman B, Harris SB, Barrett PH, Hegele RA (2002) HNF-1alpha G319S, a trans-activation-deficient mutant, is associated with altered dynamics of diabetes onset in an Oji-Cree community. *Proc Natl Acad Sci USA* 99:4614–4619
- Weiler T, Greenberg CR, Nylén E, Halliday W, Morgan K, Eggertson D, Wrogemann K (1996) Limb-girdle muscular dystrophy and Miyoshi myopathy in an aboriginal Canadian kindred map to LGMD2B and segregate with the same haplotype. *Am J Hum Genet* 59:872–878
- Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, et al (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285:901–906