Biochemical and Biophysical Research Communications **261,** 493–498 (1999)
Article ID bbrc.1999.1070, available online at http://www.idealibrary.com on IDEAL®

The Genomic Organization and Polymorphism Analysis of the Human Niemann-Pick C1 Gene

Jill A. Morris,*[,1] Dana Zhang,* Katherine G. Coleman,* James Nagle,*
Peter G. Pentchev,* and Eugene D. Carstea†

**National Institute of Neurological Disorders and Stroke, National Institutes of Health (NIH), Bethesda, Maryland 20892; and †Saccomanno Research Institute, St. Mary's Hospital and Medical Center, Grand Junction, Colorado 81502*

Received June 28, 1999

Niemann-Pick C (NP-C) is a fatal autosomal recessive storage disorder characterized by progressive neurodegeneration and variable hepatosplenomegaly. At the cellular level, cells derived from an affected individual accumulate unesterified cholesterol in lysosomes when cultured with low-density lipoprotein. The NP-C gene was identified at 18q11. The transcript is 4.9 kb encoding a 1278-amino-acid protein. We have defined the genomic structure of *NPC1* along with the 5′ flanking sequence. The *NPC1* gene spans greater than 47 kb and contains 25 exons. Exons range in size from 74 to 788 bp with introns ranging in size from 0.097 to 7 kb. All intron/exon boundaries follow the GT/AG rule. The 5′ flanking sequence has a CpG island containing multiple Sp1 sites indicative of a promoter region. The CpG island is located in the 5′ flanking sequence, exon 1 and the 5′ end of intron 1. We have also identified multiple single nucleotide polymorphisms in the coding and intronic sequences.

Niemann-Pick C disease (NP-C) is a neurovisceral lipid storage disorder (see reviews, 1–3). It has an estimated incidence as high as 1:150,000 (1). The age at which symptoms clearly appear can vary from infancy to late adulthood with the majority of patients presenting in childhood (1). The neurological profile includes progressive ataxia, dysphagia, dystonia, dementia, and the hallmark feature supranuclear vertical gaze palsy (1).

When NP-C fibroblasts are cultured with low-density lipoprotein (LDL), extensive unesterified cholesterol accumulates in lysosomes (2, 4–7). This anomalous metabolic sequestration of cholesterol has been shown to cause delays in the homeostatic responses that normally regulate cellular cholesterol levels (2, 5). Cytochemical and biochemical tests that monitor this cholesterol lipidodic state have in the past provided the only means of establishing a laboratory diagnosis of NP-C (1, 2, 5). Such cellular testing has proven difficult to interpret in cell lines that display only partial phenotypic penetrance particularly in prenatal testing and does not exist for carrier testing (8).

There are two separate disease gene loci responsible for NP-C disease. The *NPC1* gene is mutated in the majority of documented cases (~95%) and an unidentified second loci, NPC2, is estimated to account for 5% of cases (9–11). The human *NPC1* gene is located on 18q11 and has a transcript of 4.9 kb encoding a presumed protein of 1278 amino acids. Human NPC1 has 13–16 putative transmembrane domains, 12 of which have high homology to Patched, a protein involved with the Sonic hedgehog signaling pathway (9, 10, 12). Included in the twelve are five transmembrane regions with a putative sterol-sensing domains (SSD) similar to those found in SREBP Cleavage Activating Protein, SCAP, and HMG-CoA Reductase (9, 13, 14). The NPC1 protein also has an endoplasmic reticulum targeting signal sequence, a leucine zipper motif in an interspecies conserved region, and a lysosomal targeting motif.

As could be expected, the successful identification of the *NPC1* gene (9, 10) has already expedited the laboratory diagnosis of this disorder in French Acadian patients of Nova Scotia (15). A common mutation has been shown to be linked to a large number of individuals in a demographically restricted location (15). However, the initial observation of multiple separate private *NPC1* mutations in the general population has been found by several other investigators (personal communications from Dr. Marie Vanier (INSERM U.), Dr. Kosaku Ohno (Tottori U.) and Dr. Peter Bauer (U. Rostock) in March, 1999) (9). Thus the currently avail-

[1] To whom correspondence should be addressed at present address at Merck Research Laboratories, Department of Human Genetics, WP26A-3000, P.O. Box 4, West Point, PA 19486. Fax: (215) 652-2075. E-mail: jill_morris@merck.com.

Accession numbers: AF157365, AF157366, AF157367, AF157368, AF157369, AF157370, AF157371, AF157372, AF157373, AF157374, AF157375, AF157376, AF157377, AF157378, and AF157379.

0006-291X/99 $30.00

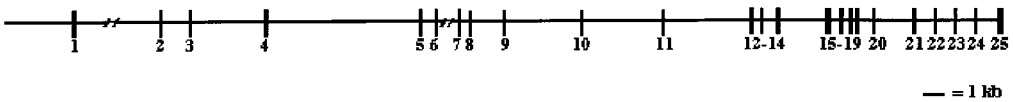

FIG. 1. The genomic structure of the *NPC1* gene. The intron sizes are to scale and the numbers below the black boxes indicate the exon number.

able technique of RT-PCR based *NPC1* mRNA analyses has its limitations in a wide spread survey of hundreds of unrelated families. The successful characterization of the full human genomic *NPC1* organization and its intron/exon sequence intervals as reported here not only expands further studies of the gene but also provides the opportunity of expediting these mutational analyses through a DNA based PCR strategy.

MATERIALS AND METHODS

Characterization of intron sizes and the intron/exon boundaries through sequencing of genomic clones and through PCR amplification from genomic DNA. The bacterial artificial chromosome (BAC) clone 108N2 (Research Genetics) was previously defined to contain *NPC1* genomic sequence (9). *Pst*I or *Bam*HI restriction enzyme digested genomic fragments from 108N2 were subcloned into pUC18 (Life Technologies). Subclones were transformed into DH5α bacteria (Life Technologies) and isolated colonies were grown in Luria-Bertani (LB) with ampicillin at 37°C from which 15% glycerol stocks were made. Glycerol stocks of individual subclones were arrayed into three 96-well plates. The 96-well glycerol stocks were replica plated onto nylon filters. The filters were laid onto LB agar plates containing ampicillin and placed at 37°C overnight. The colonies on the filters were lysed and hybridized against portions of the *NPC1* cDNA by standard techniques (16). DNA was isolated from positive colonies using the Wizard kit from Promega. The DNA was sequenced (Applied Biosystems). Seqwright also sequenced portions of the BAC 108N2 by shotgun sequencing.

Many of the intron/exon boundaries and the intron sizes were determined by long range PCR amplification of genomic DNA from unaffected individuals using *NPC1* specific primer (Clontech, Boehringer-Mannheim). PCR products were subcloned using the TA Cloning vector (Invitrogen) and sequenced multiple times in both directions across the intron/exon boundaries. Introns 5, 7, 12, 13, 15–19, 22, and 23 were completely sequenced. The sizes of the remaining introns were determined by gel electrophoresis.

Intron/exon boundaries were defined by comparing the genomic *NPC1* sequence versus the *NPC1* cDNA sequence using the MacVector program, Assemblign (Oxford Molecular). Consensus sequences were aligned using Sequencher 3.0 (Gene Codes Co.). All boundaries were sequenced with multiple reads in both directions and were sequenced at least 100 bp into the intron. Sequences were submitted to GenBank and the accession numbers are indicated (Table 1).

Characterization of the transcription start site. 5′ RACE (rapid amplification of cDNA ends) was done according to manufacturer's instruction (Life Technologies) on RNA from fibroblasts obtained from unaffected individuals with a *NPC1* specific primer. Secondly, the 5′ end of the NPC1 cDNA was PCR amplified from Marathon ready cDNA libraries from ovary and fetal tissue (Clontech). Products were analyzed by gel electrophoresis. The largest products were gel purified and sequenced (Applied Biosystems).

Characterization of the 5′ flanking sequence. By hybridization, two 108N2 BAC subclones were identified which contained the 5′ flanking sequence. The subclones were sequenced multiple times in both directions. The 5′ flanking sequence was analyzed for CpG islands using Grail version 1.3 (17). MatInspector was used to identify potential transcription factor binding sites (18). Binding sites listed have a core similarity of 1.00 and matrix similarity of greater than 0.900.

Polymorphism analysis of the NPC1 genomic sequence. Single nucleotide polymorphisms (SNPs) were identified by comparing PCR amplified *NPC1* sequences and BAC 108N2 sequences (described above) with that of the published NPC1 cDNA sequence (AF002020) (9). Polymorphisms in the intronic sequences were defined by comparing *NPC1* sequences amplified from the DNA of unaffected individuals. The recommendations of the Nomenclature Working Group were used to name the polymorphisms (19) (Table 2).

RESULTS AND DISCUSSION

The structure of the *NPC1* gene was characterized through both the sequencing of PCR products amplified from genomic DNA with *NPC1* cDNA specific primers and through the sequencing of subcloned BAC DNA. The BAC clone 108N2 was defined in the *NPC1* critical interval sequence by the genetic marker D18S75E and was shown to contain *NPC1* genomic sequence (9). We have found 108N2 to contain the entire genomic structure for the *NPC1* gene.

The entire genomic structure of the *NPC1* gene is greater than 47 kb (Fig. 1). There are 25 exons which range in size from 74 bp (exon 7) to 788 bp (exon 25) (Table 1). Introns range in size from 97 bp (intron 15) to 7 kb (intron 4). We were unable to amplify across introns 1 and 6 to determine their exact sizes. Sequence information for introns 1 and 6 is from BAC subclones. The complete sequences of introns 5, 7, 12, 13, 15–19, 22, and 23 were determined. Introns 14 and 17 vary in size by a few base pairs due to polymorphisms in repeats of Ts. The rest of the boundaries were sequenced at least 100 bp into the intron. All boundaries were analyzed with multiple sequence reads in both directions. The intron/exon boundaries all follow the GT/AG rule, GT is the splice donor and AG is the splice acceptor (20).

To define the transcription start site of *NPC1*, 5′ RACE was performed on RNA from fibroblasts obtained from unaffected individuals with a primer from the 5′ end of the *NPC1* cDNA. In addition, PCR amplification of the 5′ end of the *NPC1* cDNA was also done using Marathon ready cDNA libraries. The DNA sequence of these products did not result in any transcripts with sequences longer than the previously published transcription start site at −121 bp from the ATG

TABLE 1

The Intron/Exon Boundaries and Genomic Structure of the *NPC1* Gene

Exon	Exon size in bp[a]	Position in message[b]	Intron size in kb[c,d]	Splice acceptor site	Exon boundary sequence	Splice donor site	Accession numbers
1		1–57	>4 kb		...GCGCAG	gtgagcggtcgccggccaccgggac	AF157365
2	123	58–180	1.3	tgaagtttgtgttgactttctgtag	GTGTTT...GTGCAG	gtaagttcattatcttaggcactgg	AF157366
3	107	181–287	3.5	gtgtctgtgctctttcacctgaaag	GAACTC...GTCCAG	gtaggttctgctggggaaacagaac	AF157367
4	176	288–463	7	agaatataatttacttttcctttag	ATGTCC...CCAATG	gtaagtaaacttttaattattcctc	AF157368
5	168	464–631	**0.879**	ttttcctttgggtttcctcttctag	CAATGT...TTTCAG	gtaggtataaagattccaagtttgg	AF157369
6	250	632–881	>1	tactcaacacaattcctttctgtag	ATTTTC...CTACAG	gtaagcagttttgtttgtcatccag	AF157370
7	74	882–955	**0.503**	ctatttcttcacttctgtttttcag	AAAACG...ACAAAG	gtaggcatatttgtccatgaataaa	AF157371
8	371	956–1326	1.5	ttcccttcctctcccgttcttccag	GAGAGG...CACCAG	gtaacctgctgtttgaagaaataag	AF157371
9	227	1327–1553	3.5	gcttatttttaaaatcttctttcag	GTTCTT...CGTACG	gtaagtggcaagagacaatcattgg	AF157372
10	101	1554–1654	3.8	aggtgttttctaaacttttttggcag	GGCTCC...ATGATG	gtaagtaagagaagctttgactttt	AF157373
11	103	1655–1757	3.5	ttgatgtcacatttttcctttctag	ATCAAA...AAAAGA	gtgagtcactcatgggtgtgagcag	AF157374
12	190	1758–1947	**0.433**	ata(t/c)aacattttgcttacttttag	GTTTAT...CTTCTG	gtaagccgggggggggtggtgtatg	AF157375
13	183	1948–2130	**0.771–0.773**	tgattgtgtctgtcgcctctctcag	GTGGAT...TACCAG	gtatacttccatattctcacagggc	AF157375
14	115	2131–2245	2.2	cataa($t_{(15-17)}$)aag	AGAGAT...TCTTAG	gtaattatgctttcaatcctaccag	AF157375
15	128	2246–2373	**0.097**	c(g/a)cagagcccttctctccccgacag	GAGCAT...CAAGAG	gtaagttggtgccaggattatagtc	AF157376
16	141	2374–2514	**0.530**	cctttgtttccctctgtcccctcag	AAAAAT...ATTGTG	gtatgcgcttatctgtggtttttct	AF157376
17	90	2515–2604	**0.444–0.447**	tcttatgattcatttgttgttccag	ATAGCA...CCAGAT	gtaagatgacttcctttttttttt	AF157376
18	191	2605–2795	**0.340**	tcctattcttttatctttctttcag	GACTCC...CAACTA	gtcagtaccaccttgtcttgtactt	AF157376
19	116	2796–2911	**0.683**	cagtaaccctgtctgtcctctctag	TACCCG...CTTCAG	gtaccttctcctttaccaaatcttt	AF157376
20	130	2912–3041	1.8	agtcctccctgcatgtctccgccag	TGGTTG...CAAAGG	gtaagtgctgctgccattgcagata	AF157376
21	204	3042–3245	1.0	ttgacacccaggattctttcctcag	GGGACA...TTACAG	gtaaagcctgccctttttcaatggg	AF157377
22	232	3246–3477	**0.090**	atgaggcctcccctctcccctccag	TGTGTT...GTGATG	gtgagtcctcatacagtctcagttt	AF157378
23	114	3478–3591	**0.928**	tcctaaaggaagtgctttatttcag	AGCTGT...AGCTCC	gtgagtaccctgagcagggccacgc	AF157378
24	163	3592–3754	1.1	ttctaacacagtatctcttcttttag	GTGTTC...ACATAG	gtaagagttctcatcttaaaagggt	AF157378
25	788	3755–4542		aagaaaaagtctctctctctcatag	GGCCAT...		AF157379

[a] Base pair(s).
[b] The nucleotide positions of the exons in the *NPC1* message are indicated with the A of ATG being +1.
[c] Kilobase(s).
[d] Intron sizes indicated in bold are introns in which the entire intron was sequenced. The remaining introns were sized by gel electrophoresis.

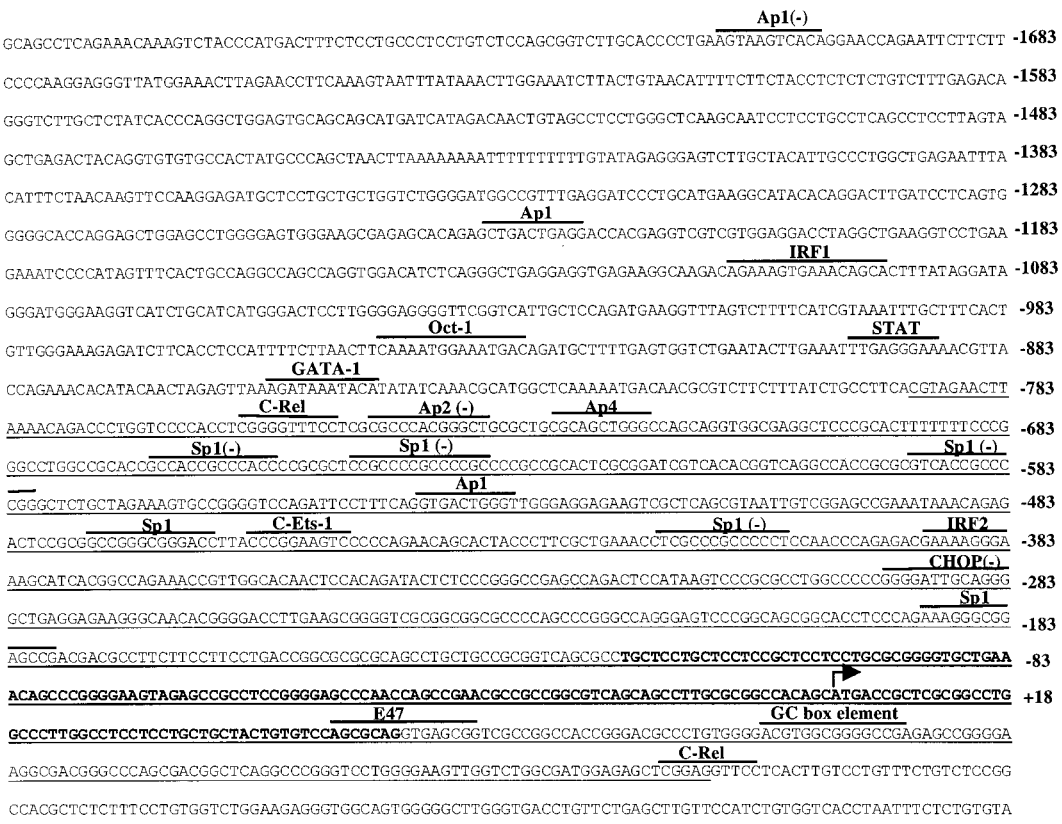

FIG. 2. The 5′ flanking sequence of *NPC1* gene. The bold type indicates transcribed sequence. The underlined sequence is the CpG island determined by Grail version 1.3 (17). Potential transcription factor binding sites were determined by MatInspector (core similarity of 1.00 and matrix similarity of >0.900) (18). Binding sites with a (−) are in the antisense direction. The complete 3-kb sequence is deposited in GenBank (AF157365).

(the A being nucleotide +1) (9). This data suggests that the genomic sequence upstream of the transcription start site (−121) contains the *NPC1* promoter region.

Two genomic subclones from the BAC 108N2 contained the 5′ flanking sequence of the *NPC1* gene (Fig. 2). One clone (*Bam*HI G5) contained 3 kb of 5′ flanking sequence along with exon 1 and the 5′ end of intron 1. The 5′ flanking region does not contain a TATA sequence, but it did contain a GC rich sequence with six Sp1 sites indicative of the promoter of the gene (21–23). CpG island analysis using Grail yielded a CpG score of 0.848 with a GC composition of 69.7% on a 980-bp region of the 5′ flanking sequence (17). CpG islands are defined by a CpG score greater than 0.6 with a percentage GC of greater than 50% (21). The CpG island extended from nucleotide −792 into the 5′ flanking sequence (A of ATG being +1) through exon 1 into the 5′ end of intron 1. The 5′ end of intron one has a GC rich region with 80% GC composition compared to all the other 5′ intronic sequences in the *NPC1* gene which have a GC content of 60% or less. CpG islands have been defined to be at the transcription start site of many housekeeping genes (21, 22). Their open conformations are believed to allow the binding of transcription factors to promoter regions (22, 24). MatInspector was used to define multiple potential transcription factor binding sites (18) (Fig. 2). Further experiments need to be performed to define the biological relevance of these putative binding sites.

Polymorphisms were identified in the coding and intronic sequences of the *NPC1* gene (Table 2). Ten SNPs were identified in the coding sequence by comparing the published *NPC1* cDNA sequence (AF002020) with sequences amplified from genomic DNA and by RT-PCR (Table 2a). This sequence amplification was performed on DNA and RNA from unaffected individuals. Three of the SNPs (at nucleotides 1755, 2226, and 2793) did not change the encoded amino acid. The rest of the SNPs resulted in relatively conservative changes. At nucleotide 644, there is a SNP that results in an arginine instead of a histidine, which are both basic amino acids. At nucleotide 998, a SNP was identified in unaffected individuals that results in a glycine being an aspartic acid. The mouse Npc1 ortholog also has an aspartic acid at this position. In addition, at nucleotide 2572, an A > G polymorphism results in an isoleucine being a valine. In the mouse Npc1 ortholog, there is an isoleucine at this

TABLE 2a

Single Nucleotide Polymorphisms Identified in the *NPC1* Coding Sequence

Nucleotide[a,b]	Amino acid change	Region in the NPC1 protein	Amino acids in NPC1 orthologs[c]		
			Mouse	Yeast	Worm
c644A > G	215 His → Arg	—	Leu	Glu	Thr
c998G > A	333 Gly → Asp	—	Asp	Gly	Asn
c1415T > C	472 Leu → Pro	—	Leu	Thr	Met
c1755A > G	585 Lys → Lys	—	Lys	Glu	Lys
c1926C > G	642Ile → Met	Transmembrane	Ile	—	Tyr
c2226G > A	742 Glu → Glu	—	Glu	Gln	Cys
c2270T > C	757 Val → Ala	Transmembrane	Val	Val	Ile
c2572A > G	858 Ile → Val	Transmembrane	Ile	Phe	Val
c2793C > T	931 Asn → Asn	—	Thr	His	Glu
c3659T > C	1220 Ile → Thr	—	Ile	Ile	Ile

[a] c indicates that the nucleotide position corresponds to the *NPC1* cDNA sequence with the A of ATG being +1.
[b] All polymorphisms are named according to the recommendations of the Nomenclature Working Group (19).
[c] The orthologous amino acids from mouse (AF003348), yeast (*Saccharomyces cerevisiae*, Lpa11p, U33335), and worm (*Caenorhabditis elegans*, F02E8p, U53340) NPC1 proteins were identified by ClustalW Alignments (MacVector).

position whereas the worm ortholog has a valine. One of the ten SNPs identified was from RT-PCR product from RNA from unaffected fibroblast. The SNP at nucleotide 1415 was identified in the *NPC1* cDNA 1-1 that was previously shown to be an active form of the protein (9). Twenty-six polymorphisms were identified in the intronic sequences of the *NPC1* gene (Table 2b). The polymorphisms were identified by comparing amplified sequences from unaffected individuals. Recently, there has been a great interest in SNPs for their usefulness in genome scans and association studies (25).

TABLE 2b

Polymorphisms in the Intronic Sequence of the *NPC1* Gene

Intron	Polymorphism[a,b]	Flanking sequences
6	IVS6 − 47A > C	aaacaacctc(a/c)ctgtgatgaa
10	IVS10 + 240A > G	gtcgaaggct(a/g)agtacctttc
11	IVS11 + 100G > A	aaccaatggc(g/a)gatctgagtt
11	IVS11 − 178A > G	ccgggaggtg(a/g)agcttggagt
11	IVS11 − 22T > C	actaaaaata(t/c)aacattttgc
13	IVS13+ 433A > G	gctgctgtag(a/g)aggtggtctc
13	IVS13 + 570T > A	gctttccctt(t/a)agtttctcac
13	IVS13 + 591G > A	ttggtaccta(g/a)ctctgtagtg
13	IVS13 + 754delT	tgtcgcataa(delt)tttttttttt
13	IVS13 + 754 − 755delTT	tgtcgcataa(deltt)ttttttttttt
14	IVS14 + 95delT	tgttgcttcc(delt)tttttttttt
14	IVS14 + 121 − 122delTT	ttttttttt(deltt)ctctgagacgg
14	IVS14 − 24G > A	cagaagtgac(g/a)cagagccctt
16	IVS16 + 163C > T	atgtttttcc(c/t)gtgagctttt
17	IVS17 + 15delT	gatgacttcc(delt)tttttttttt
17	IVS17 + 15 − 17delTTT	gatgacttcc(delttt)ttttttttttt
17	IVS17 + 378A > G	cttattctcc(a/g)tgatcctcgc
19	IVS19 + 28T > C	aaatctttcc(t/c)gttttgctga
20	IVS20 − 158T > C	aaacctttgg(t/c)gttcccttat
21	IVS21 − 306delT	ttgttttttt(delt)gtgttttttc
21	IVS21-280G > A	gctgtggaat(g/a)ctctctgtaa
24	IVS24 + 75T > A	tgaggcattc(t/a)tttcaagtat
24	IVS24 − 123C > T	gttgtccagg(c/t)tggtctcaag
24	IVS24 − 89A > G	tcagtcttcc(a/g)aagtgggatt
24	IVS24 − 75G > A	tgggattaca(g/a)gcgtgagcca

[a] IVS stands for intervening sequence.
[b] All polymorphisms are named according to the recommendations of the Nomenclature Working Group (19).

CONCLUSIONS

The characterization of the *NPC1* genomic structure will be invaluable for mutation detection in this pan-ethnic disorder. It will allow for the diagnosis of patients by genomic DNA as opposed to the currently available methods of RT-PCR or biochemical testing (1, 2, 8). In this report, we described the 25 intron/exon boundaries. In addition, the putative promoter sequence which contains a CpG island and multiple Sp1 sites, indicative of a promoter region, was identified (22, 23). Thirty-six polymorphisms were also determined.

ACKNOWLEDGMENTS

The Ara Parseghian Medical Research Foundation and the National Niemann-Pick Disease Foundation supported this work.

REFERENCES

1. Patterson, M. C., Vanier, M. T., Suzuki, K., Morris, J. A., Carstea, E. D., Neufeld, E., Blanchett-Mackie, E. J., and Pentchev, P. G. (1997) *in* The Metabolic and Molecular Bases of Inherited Disease (Scriver, C. R., Sly, W. S., Beaudet, A. L., and Valle, D., Eds.), McGraw Hill, New York.
2. Pentchev, P. G., Brady, R. O., Blanchette-Mackie, E. J., Vanier, M. T., Carstea, E. D., Parker, C. C., Goldin, E., and Roff, C. F. (1994) *Biochim. Biophys. Acta* **1225,** 235–243.
3. Morris, J. A., and Carstea, E. D. (1998) *Mol. Med. Today* **4,** 525–531.
4. Pentchev, P. G., Comly, M. E., Kruth, H. S., Patel, S., Proestel, M., and Weintroub, H. (1986) *J. Biol. Chem.* **261,** 2772–2777.
5. Pentchev, P. G., Comly, M. E., Kruth, H. S., Tokoro, T., Butler, J., Sokol, J., Filling-Katz, M., Quirk, J. M., Marshall, D. C., Patel, S., Vanier, M. T., and Brady, R. O. (1987) *FASEB J.* **1,** 40–45.
6. Sokol, J., Blanchette-Mackie, E. J., Kruth, H. S., Dwyer, N. K., Amende, L. M., Butler, J. D., Robinson, E., Patel, S., Brady, R. O., Comly, M. E., Vanier, M. T., and Pentchev, P. G. (1988) *J. Biol. Chem.* **263,** 3411–3417.
7. Neufeld, E. B., Wastney, M., Patel, S., Suresh, S., Cooney, A. M., Dwyer, N. K., Roff, C. F., Ohno, K., Morris, J. A., Carstea, E. D., Incardona, J. P., Strauss, J. F., III, Vanier, M. T., Patterson, M. C., Brady, R. O., Pentchev, P. G., and Blanchette-Mackie, E. J. (1999) *J. Biol. Chem.* **274,** 9627–9635.
8. Vanier, T. M., Rodriguez-Lafrasse, C., Rousson, R., Mandon, G., Boue, J., Choiset, A., Peyrat, M., Dumontel, C., Juge M., Pentchev, P. G., Revol, A., and Louisot, P. (1992) *Am. J. Hum. Genet.* **51,** 111–122.
9. Carstea, E. D., Morris, J. A., Coleman, K. G., Loftus, S. K., Zhang, D., Cummings, C., Gu, J., Rosenfeld, M. A., Pavan, W. J., Krizman, D. B., Nagle, J., Polymeropoulos, M. H., Sturley, S. L., Ioannou, Y. A., Higgins, M. E., Comly, M., Cooney, A., Brown, A., Kaneski, C. R., Blanchette-Mackie, E. J., Dwyer, N. K., Neufeld, E. B., Chang, T., Liscum, L., Strauss, J. F., Ohno, K., Zeigler, M., Carmi, R., Sokol, J., Markie, D., O'Neill, R. R., Van Diggelen, O. P., Elleder, M., Patterson, M. C., Brady, R. O., Vanier, M. T., Pentchev, P. G., and Tagle, D. A. (1997) *Science* **277,** 228–231.
10. Loftus, S. K., Morris, J. A., Carstea, E. D., Gu, J. Z., Cummings, C., Brown, A., Ellison, J., Ohno, K., Rosenfeld, M. A., Tagle, D. A., Pentchev, P. G., and Pavan, W. J. (1997) *Science* **277,** 232–235.
11. Vanier, M. T., Duthel, S., Rodriguez-Lafrasse, C., Pentchev, P., and Carstea, E. D. (1996) *Am. J. Hum. Genet.* **58,** 118–125.
12. Johnson, R. L., Rothman, A. L., Xie, J., Goodrich, L. V., Bare, J. W., Bonifas, J. M., Quinn, A. G., Myers, R. M., Cox, D. R., Epstein, E. H., Jr., and Scott, M. P. (1996) *Science* **272,** 1668–1671.
13. Chin, D. J., Gil, G., Russell, D. W., Liscum, L., Luskey, K. L., Basu, S. K., Okayama, H., Berg, P., Goldstein, J. L., and Brown, M. S. (1984) *Nature* **308,** 613–617.
14. Hua, X., Nohturfft, A., Goldstein, J. L., and Brown, M. S. (1996) *Cell* **87,** 415–426.
15. Greer, W. L., Riddell, D. C., Gillan, T. L., Girouard, G. S., Sparrow, S. M., Byers, D. M., Dobson, M. J., and Neumann, P. E. (1998) *Am. J. Hum. Genet.* **63,** 52–54.
16. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York.
17. Uberbacher, E. C., Xu, Y., and Mural, R. J. (1996) *Methods Enzymol.* **266,** 259–281.
18. Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. (1995) *Nucleic Acids Res.* **23,** 4878–4884.
19. Antonarakis, S. E., and the Nomenclature Working Group. (1998) *Hum. Mut.* **11,** 1–3.
20. Breathnach, R., and Chambon, P. (1981) *in* Annual Review of Biochemistry (Snell, E. E., Boyer, P. D., Meister, A., and Richardson, C. C., Eds.), pp. 349–383, Annual Reviews, CA.
21. Gardiner-Garden, M., and Frommer, M. (1987) *J. Mol. Biol.* **196,** 261–282.
22. Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992) *Genomics* **13,** 1095–1107.
23. McKnight, S., and Tijan, R. (1986) *Cell* **46,** 795–805.
24. Tazi, J., and Bird, A. (1990) *Cell* **60,** 909–920.
25. Collins, F. S., Brooks, L. D., and Chakravarti, A. (1998) *Genome Res.* **8,** 1229–1231.