Cloning and Characterization of the Mouse Mcoln1 Gene Reveals an Alternatively Spliced Transcript Not Seen in Humans

BMC Genomics

BioMed Central

Research article

Cloning and characterization of the mouse *Mcoln1* gene reveals an alternatively spliced transcript not seen in humans

John L Falardeau†[1,3], John C Kennedy†[1,3], James S Acierno Jr[1,3], Mei Sun[2], Stefanie Stahl[2], Ehud Goldin[2] and Susan A Slaugenhaupt*[1,3]

Address: [1]Harvard Institute of Human Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA, [2]Developmental and Metabolic Neurology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA and [3]Molecular Neurogenetics Unit, Massachusetts General Hospital, Charlestown, MA, USA

E-mail: John L Falardeau - falardea@helix.mgh.harvard.edu; John C Kennedy - jkennedyhelix@yahoo.com; James S Acierno - acierno@helix.mgh.harvard.edu; Mei Sun - msun@cpdr.org; Stefanie Stahl - stahl@codon.nih.gov; Ehud Goldin - goldin@codon.nih.gov; Susan A Slaugenhaupt* - slaugenh@helix.mgh.harvard.edu

*Corresponding author †Equal contributors

Published: 5 February 2002

BMC Genomics 2002, **3**:3

Received: 27 November 2001
Accepted: 5 February 2002

This article is available from: http://www.biomedcentral.com/1471-2164/3/3

Abstract

Background: Mucolipidosis type IV (MLIV) is an autosomal recessive lysosomal storage disorder characterized by severe neurologic and ophthalmologic abnormalities. Recently the MLIV gene, *MCOLN1*, has been identified as a new member of the transient receptor potential (TRP) cation channel superfamily. Here we report the cloning and characterization of the mouse homologue, *Mcoln1*, and report a novel splice variant that is not seen in humans.

Results: The human and mouse genes display a high degree of synteny. *Mcoln1* shows 91% amino acid and 86% nucleotide identity to *MCOLN1*. Also, *Mcoln1* maps to chromosome 8 and contains an open reading frame of 580 amino acids, with a transcript length of approximately 2 kb encoded by 14 exons, similar to its human counterpart. The transcript that results from murine specific alternative splicing encodes a 611 amino acid protein that differs at the c-terminus.

Conclusions: *Mcoln1* is highly similar to *MCOLN1*, especially in the transmembrane domains and ion pore region. Also, the late endosomal/lysosomal targeting signal is conserved, supporting the hypothesis that the protein is localized to these vesicle membranes. To date, there are very few reports describing species-specific splice variants. While identification of *Mcoln1* is crucial to the development of mouse models for MLIV, the fact that there are two transcripts in mice suggests an additional or alternate function of the gene that may complicate phenotypic assessment.

Background

Mucolipidosis type IV (MLIV; MIM 252650) is an autosomal recessive lysosomal storage disorder that is characterized by corneal clouding, delayed psychomotor development, and mental retardation that usually presents during the first year of life [1]. Another interesting clinical characteristic is that patients are constitutively achlorhydric with associated hypergastremia [2]. Patients with MLIV do not show mucopolysaccharide excretion, skeletal changes, or organomegaly like the other mucolipidoses. Abnormal lysosomal storage bodies and large vac-

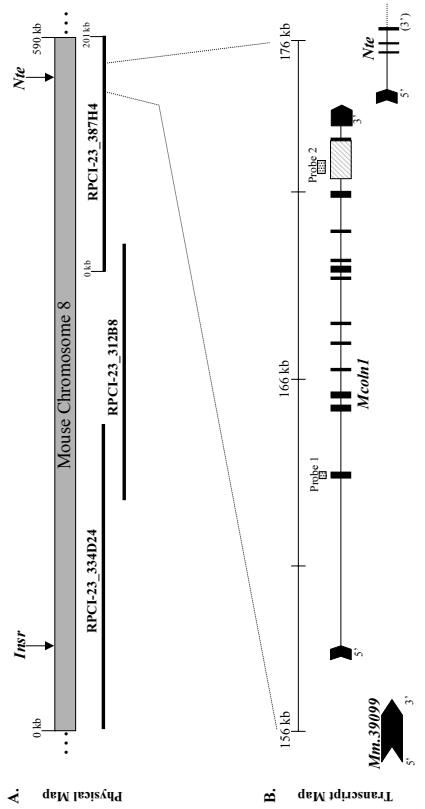

Figure 1
Physical and transcript map of the *Mcoln1* gene region. (**A**) Physical map showing a 590 kb segment of mouse chromosome 8 syntenous to human chromosome 19, anchored by the genes *Insr* and *Nte*. The region is covered by three BACs: RPCI-23_334D24 (AC087153), RPCI-23_312B8 (AC087150), and RPCI-23_387H4 (AC079544). (**B**) The transcript map of *Mcoln1* shows 14 exons, and the locations of the probes used in this study are illustrated. The map also shows the expanded exon 13 in the alternatively spliced transcript (hashed box). UniGene cluster Mm.39099 is the homologue of the human zinc finger gene (AC001252). It should be noted that the scale of the transcript map is in reference to the 201 kb scale of BAC RPCI-23_387H4.

uoles have been found in skin and conjuctival biopsies using electron-microscopy and, prior to gene identification, served as the only means of diagnosis [3–5]. A recent report estimates that the carrier frequency of MLIV in the Ashkenazi Jewish population is 1 in 100, and mutations have been reported in Jewish and non-Jewish families [6–9].

The human gene *MCOLN1* (GenBank #AF287270) maps to chromosome 19p13.2-13.3 and encodes a novel protein that is a member of the transient receptor potential (TRP) cation channel gene superfamily [7–10]. Protein trafficking studies suggest that MLIV is the result of a defect in the late endocytic pathway, contrary to the other mucolipidoses which are typically caused by defective lysosomal hydrolases [11,12]. Recent work in *Caenorhabditis elegans* supports this hypothesis. Loss of function mutants of the *MCOLN1 C. elegans* homologue, *cup-5*, result in an increased rate of endocytosis, accumulation of large vacuoles, and a decreased rate of endocytosed protein breakdown; while over-expression of this gene reverses the phenotype [13]. Cloning and characterization of the mouse homologue of *MCOLN1* is crucial for the development of mouse models of MLIV to further study this disorder.

Results and discussion

Cloning and mapping of the mouse homologue Mcoln1

In order to clone the mouse homologue of *MCOLN1*, the human amino acid sequence was compared to the high throughput genomic sequence (HTGS) database using TBLASTN, which identified the mouse BAC clone RPCI-23_387H4 (GB No. AC079544.1). Correspondence with the Joint Genome Institute and the Lawrence Livermore National Laboratory (LLNL) Human Genome Center confirmed the location of this BAC to mouse chromosome 8 and allowed us to construct a physical map of this region [14] (Fig. 1A). The BAC sequence was then compared to the mouse EST database using BLASTN, and multiple ESTs and their corresponding I.M.A.G.E. clones were identified.

Three clones (ID Nos. 604971, 1228665, 1247566) from the UniGene cluster Mm. 8356 were obtained, sequenced, and assembled into a 2 kb transcript with an open reading frame of 580 amino acids designated *Mcoln1* (GB No. AF302010). Comparison of the cDNA sequence to the BAC clone showed that the gene consists of 14 exons. We then created a transcript map of the *Mcoln1* region of mouse chromosome 8 (Fig. 1B), noting the presence of the UniGene Cluster Mm. 39099, a homologue of the human zinc finger gene (GB No. AC001252) that terminates approximately 1.8 kb before the start of *Mcoln1*; and *Nte*, the mouse homologue of the human neuropathy target esterase gene (*NTE*) beginning 130 base pairs after the polyadenylation signal for *Mcoln1*. The transcript map of the region surrounding *Mcoln1* is similar to the corresponding region of the homologous human *MCOLN1* region [15], and the presence of the zinc finger gene and *Nte* confirms and extends the region of synteny between human chromosome 19 and mouse chromosome 8.

Characterization of Mcoln1

Comparison of the mouse and human peptide sequences showed 91% identity (Fig. 2). The *C. elegans* homologue *cup-5* shows 34% identity with *Mcoln1* and BLASTP analysis of *Mcoln1* identified a putative *Drosophila melanogaster* homologue that shows 38% identity (Fig. 2). Interestingly, two *MCOLN1* amino acid substitutions that result in

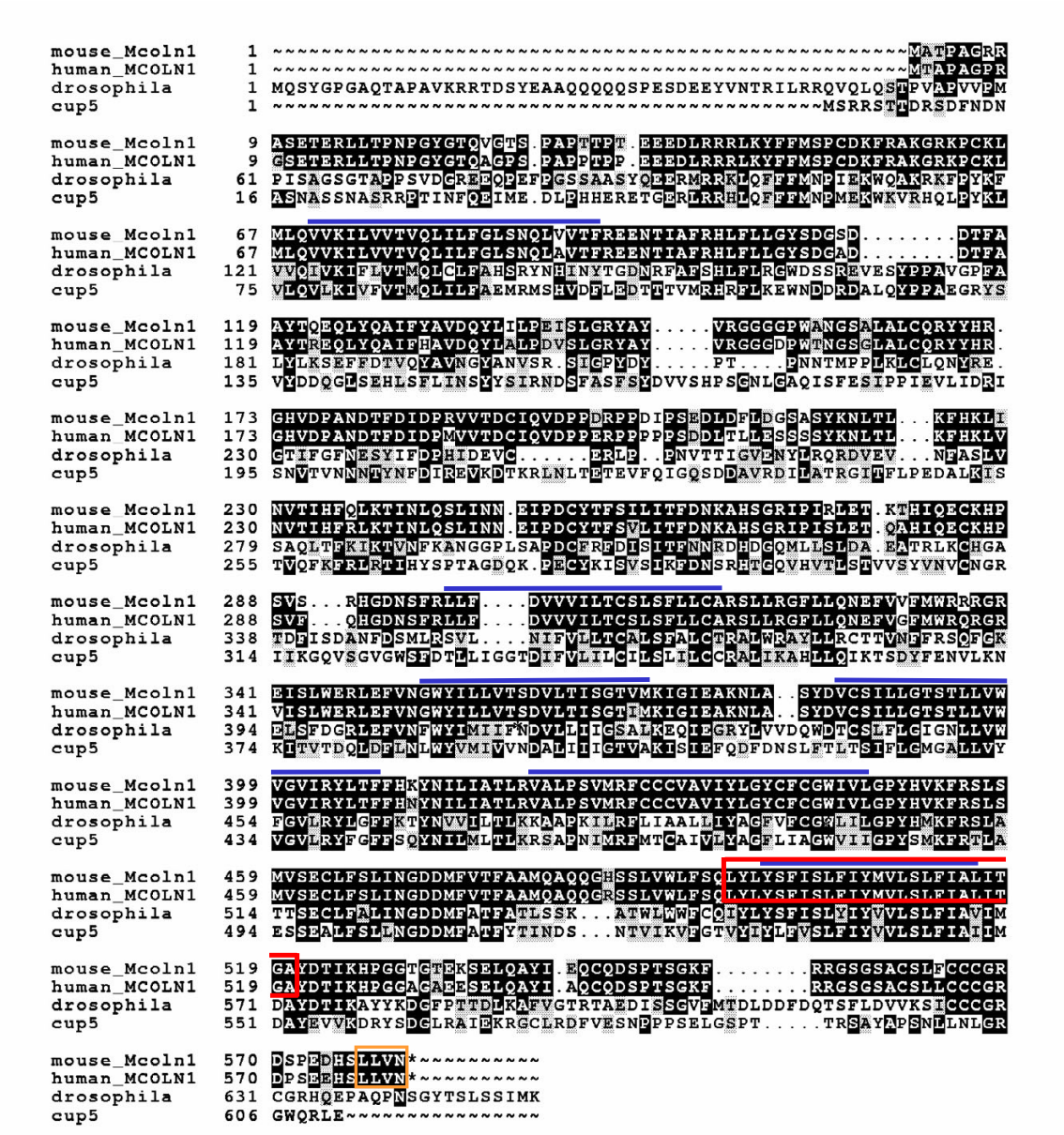

Figure 2
Peptide sequence comparison of *Mcoln1* to human, *D. melanogaster*, and *C. elegans (cup-5)* homologues. Blue lines indicate transmembrane domains, the red box surrounds the putative channel pore, the orange box surrounds the putative late endosomal/lysosomal targeting signal.

MLIV occur at conserved amino acids. TMPred analysis [http://www.ch.embnet.org/software/TMPRED_form.html] predicts a protein structure that is nearly identical to *MCOLN1*, containing 6 transmembrane domains with the N- and C-termini residing in the cytoplasm (Fig. 2) [9].

Expression analysis of Mcoln1

Mouse adult multiple tissue and embryonic Northern blots were hybridized using a probe generated from mouse exon 2 (probe 1, Fig. 1B), yielding a band of approximately 2.4 kb (isoform 1), as expected, and a less abundant and unexpected 4.4 kb band (isoform 2) (Figs. 3A & 3B). The 2.4 kb band shows ubiquitous but variable expression, with the highest expression in brain, liver and kidney. The fetal tissue blot shows decreasing levels of the 2.4 kb message with increasing gestational age. Since the human MCOLN1 gene encodes a single transcript [7–9], we carried out additional hybridizations with probes generated from various regions of the coding sequence and 3'

UTR, and all probes identified the same two transcripts in the mouse (data not shown). In order to verify the presence of a single mouse locus, we hybridized a mouse Southern blot with the exon 2 probe. Four different restriction enzymes were used, and only the expected size bands for the chromosome 8 locus were detected (data not shown).

Characterization of the Mcoln1 alternative splice variant

In order to determine the coding sequence for the larger transcript, we searched the mouse EST database using each intron as well as the genomic sequence flanking the *Mcoln1* gene. Two ESTs were identified that contained sequence from intron 12 (GB No. AI430291 and AA874645), and the corresponding clones were sequenced. Clone 408619 (ESTs: GB No. AI430291, AI429558) begins approximately 1.1 kb before exon 13 and continues through the exon and splices correctly to exon 14. Clone 1281641 (EST:GB No. AA874645) begins 175 bp before exon 13 and also splices correctly to exon 14. A mouse multiple tissue Northern was hybridized using a probe generated from the putative intron sequence in clone 408619 (Probe 2, Fig. 1B), which detected only the 4.4 kb band (Fig. 3C).

In order to determine the sequence of the entire transcript, RT-PCR using primers in exons 10 and 11 paired with a primer in intron 12 was performed using BALB/c mouse brain total RNA and the resulting products sequenced. These products show that the larger transcript is due to an alternative splice event that results in an expanded exon 13. Specifically, exon 12 splices at bp 436 of intron 12, creating a large 1614 bp exon 13 that splices correctly to exon 14. The open reading frame of this alternatively spliced transcript is 611 amino acids, 28 amino acids longer than the message encoded by the 2.4 kb transcript.

TMPred analysis predicts that isoform 2 encodes a protein identical in structure to *Mcoln1*, possessing 6 transmembrane domains and a channel pore, however the protein sequences diverge at amino acid 526. The 55 amino acid C-terminal cytoplasmic tail encoded by the 2.4 kb transcript is completely different from the 86 amino acid tail encoded by the murine specific 4.4 kb transcript (Fig. 4). Clontech Mouse RNA Master Blots were hybridized with the exon 2 and intron 12 probes mentioned above in an attempt to determine if these two transcripts showed differences in expression patterns, however, there was no significant difference in the 22 tissues represented (data not shown).

Next, we directly compared the nucleotide and amino acid sequence of the alternatively spliced mouse transcript to the entire human *MCOLN1* genomic sequence and found no significant similarity. As mentioned previously,

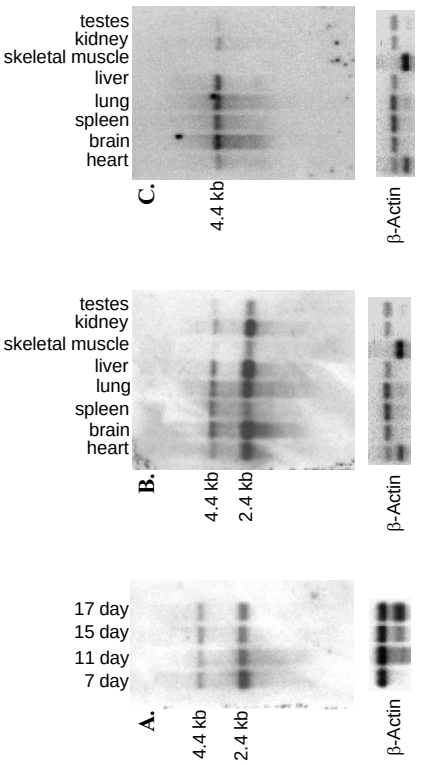

Figure 3
Northern analysis of *Mcoln1*. (**A**) Mouse fetal tissue Northern and (**B**) mouse multiple tissue Northern blots (Clontech) hybridized with a probe generated from exon 2 (probe 1, Fig. 1B). (**C**) Hybridization of the multiple tissue blot with a probe from the alternatively spliced segment of exon 13 (probe 2, Fig. 1B). β-Actin control hybridization is shown below each blot.

Northern blots performed with human *MCOLN1* probes show only one 2.4 kb transcript. In addition, we hybridized a human multiple tissue Northern and human Southern with a probe in human intron 12 that is adjacent to exon 13. The probe was located in the region syntenic to that which encodes the alternate mouse transcript. Only the expected bands were detected on the Southern and no bands were detected on the Northern, confirming that this alternative transcript is specific to murine *Mcoln1*. Recent BLASTP analysis of the alternate *Mcoln1* transcript yields a match to a putative 145 amino acid anonymous protein (GB No. BAB25862) predicted from a RIKEN clone. It is obvious from our results, however, that the identification of this sequence as a full-length protein is incorrect since probes unique to the clone, as well as probes containing the *Mcoln1* coding sequence, identify the same transcripts.

Conclusions

Comparison of *Mcoln1* isoform 1 to its human homologue shows striking similarity at both the amino acid and nucleotide level. All six of the transmembrane domains, as well as the putative cation channel are highly conserved. The putative di-leucine (L-L-X-X) motif at the C-terminus, which may act as a late endosomal/lysosomal targeting signal, is also conserved [9]. This speculation is supported by work with *cup-5*[13], the *c. elegans* homologue of *MCOLN1*, since cellular localization studies suggest that the protein is found in the late endosomes and/or lysosomes.

The mouse *Mcoln1* gene has two alternatively spliced isoforms, with isoform 2 having a different c-terminal cytoplasmic tail. The unique 86 amino acid c-terminal tail lacks the lysosomal targeting signal and does not contain any conserved domains when compared against the current profile databases. We speculate that this protein may have similar channel function but an alternate subcellular localization, but this must be proven once isoform-specific antibodies are raised. However, our results suggest that phenotypic assessment of *Mcoln1* knock-out mice may be complicated and that care must be taken when interpreting data on mouse gene expression and phenotype.

Interestingly, the second *Mcoln1* isoform is not seen in humans and the sequence of the alternatively spliced region is not conserved between man and mouse. To date, very few genes have been reported that show species specific alternative splice variants. MOG, myelin/oligodendrocyte glycoprotein, has many different splice variants in humans that are not found in mice [17]. ATP11B, a P-type ATPase, has a rabbit-specific splice variant that deletes a transmembrane domain and therefore likely alters the putative function of the protein [18]. Sequencing of the human genome has led to estimates of approximately 32,000 genes, a total surprise given the previous significantly higher estimates that were based on the number of expressed sequence tags (ESTs) in the public databases. This apparent disparity suggests a major role for alternative splicing in creating genetic complexity, and has brought the study of splicing regulation to the forefront of molecular genetics. It is likely that an abundance of species-specific splice variants will be identified as the characterization of alternatively spliced transcripts progresses.

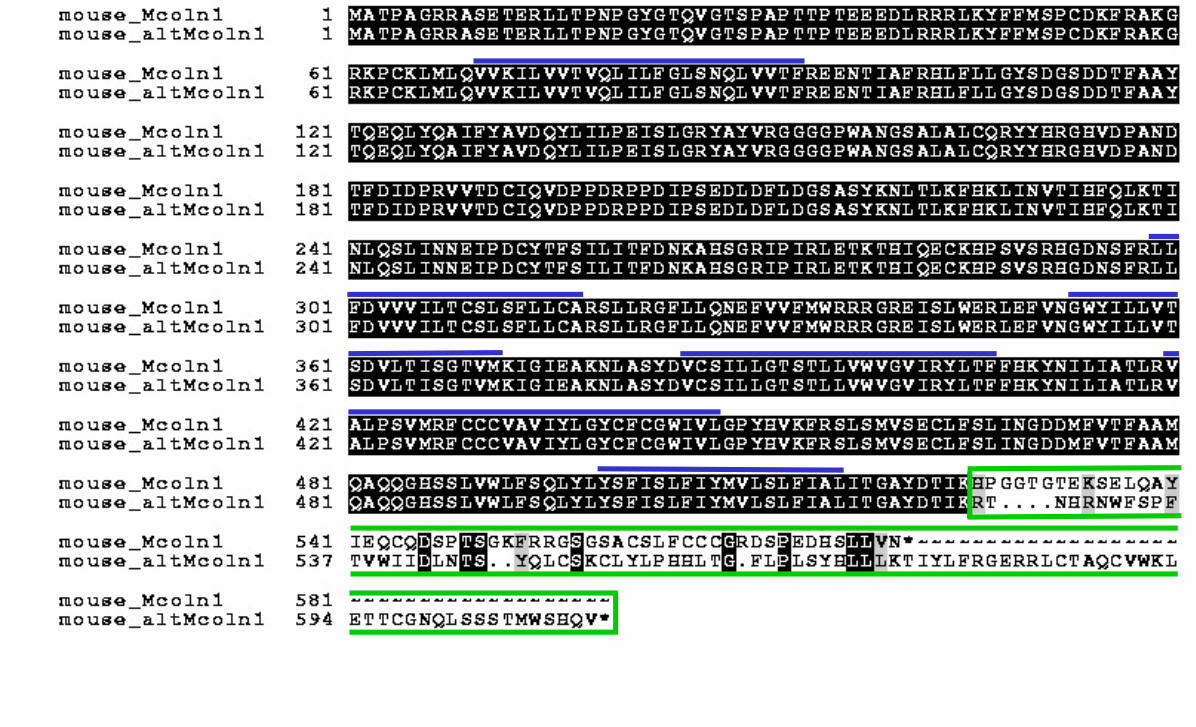

Figure 4
Peptide sequence comparison of the two alternatively spliced *Mcoln1* isoforms. The green box surrounds the divergent c-terminal cytoplasmic tails. The blue lines indicate the transmembrane domains.

Materials and methods

Bioinformatics

We conducted database searches using BLAST [http://www.ncbi.nlm.nih.gov/blast] and Draft Human Genome Browser [http://genome.ucsc.edu/] . Sequences from UniGene [http://www.ncbi.nlm.nih.gov/UniGene] were used to confirm the *Mcoln1* sequence. We performed motif searches using ProfileScan [http://www.isrec.isb-sib.ch/software/PFSCAN_form.html] and TMPred [http://www.ch.embnet.org/software/TMPRED_form.html] and alignment of protein sequences using Pileup (GCG) and Boxshade [http://www.ch.embnet.org/softward/BOX_form.html] .

DNA Sequencing

I.M.A.G.E. Clones were purchased from Research Genetics (Huntsville, AL). Sequencing was performed using the AmpliCycle sequencing kit from Applied Biosystems (Foster City, CA) on a Genomyx LR programmable DNA sequencer.

RT-PCR Reaction

BALB/c mouse brain total RNA was purchased from Clontech Laboratories (Palo Alto, CA) and made to a 1 μg / ml concentration. This RNA was used as a template to create cDNA via RT-PCR using random hexamers and oligo dT primer. The RT product was used to confirm the alternatively spliced form using primers: 5'-CATCTACCTGGGCTATTGC-3' (forward) and 5'-GCTCTCAGGTGGTGGACAC-3' (reverse) in a PCR reaction with an annealing temperature of 61°C.

Southern Blot Analysis

Total mouse genomic DNA was digested using *EcoRI, BamHI, PstI,* and *XbaI*. The digests were electrophoresed on a 1% agarose gel at 60V overnight and were transferred onto a Hybond N+ membrane from Amersham Pharmacia Biotech (Piscataway, NJ). A ^{32}P-dATP labeled PCR fragment of the *Mcoln1* coding region corresponding to exon 2 was used as a probe (primers 5'-CCCCACAGAAGAGGAAGAC-3' (forward) and 5'-AGATCTTGACCACCTGCAG-3' (reverse) with an annealing temperature of 59°C). Hybridization and washes were carried out in standard conditions [16].

Northern Blot Analysis

Mouse embryo multiple-tissue northern blot and mouse adult multiple-tissue northern blot filters were purchased from Clontech Laboratories (Palo Alto, CA). The filters were hybridized with the ^{32}P-dATP labeled DNA fragment of *Mcoln1* coding region corresponding to exon 2 (see above). For the alternative transcript, a probe was generated in the region between exons 12 and 13 with primers 5'-GTGTCCACCACCTGAGAG-3' (forward) and 5'-GAAGTAGCATTCCTGCAGGC-3' (reverse) with an annealing temperature of 62°C. The filters were then hybridized with β-actin probes. Hybridizations and washes were carried out in standard conditions, with the stripping of previously bound probes in between [16].

Acknowledgements

This work was supported by grant NS39995.

References

1. Amir N, Zlotogora J, Bach G: **Mucolipidosis Type IV: Clinical Spectrum and Natural History.** *Pediatrics* 1987, **79:**953-959
2. Schiffmann R, Dwyer NK, Lubensky IA, Tsokos M, Sutliff VE, Latimer JS, Frei KP, Brady RO, Barton NW, Blanchette-Mackie EJ, Goldin E: **Constitive achlorohydria in mucolipidosis type IV.** *Proc Natl Acad Sci USA* 1998, **95:**1207-1212
3. Berman ER, Livni N, Shapira E, Merin S, Levij IS: **Congenital corneal clouding with abnormal systemic storage bodies: a new variant of Mucolipidosis.** *J Pediatr* 1974, **84:**519-526
4. Crandall BF, Philippart M, Brown WJ, Blustone DA: **Review article: Mucolipidosis IV.** *Am J Med Genet* 1982, **12:**301-308
5. Tellez-Nagel I, Rapin I, Iwamoto T, Johnson AB, Norton WT, Nitowsky H: **Mucolipidosis IV. Clinical, ultrastructural, histochemical, and chemical studies of a case, including a brain biopsy.** *Arch Neurol* 1976, **33:**828-835
6. Bargal R, Avidan N, Olender T, Asher EB, Zeigler M, Raas-Rothschild A, Frumkin A, Ben-Yoseph O, Friendlender Y, Lancet D, Bach G: **Mucolipidosis Type IV: Novel *MCOLN1* Mutations in Jewish and Non-Jewish Patients of the Disease in the Ashkenazi Jewish Population.** *Hum Mutat* 2001, **17:**397-402
7. Bargal R, Avidan N, Ben-Asher E, Olender Z, Zeigler M, Frumkin A, Rass-Rothschild A, Glusman G, Lancet D, Bach G: **Identification of the gene causing Mucolipidosis type IV.** *Nat Genet* 2000, **26:**118-123
8. Bassi MT, Manzoni M, Monti E, Pizzo MT, Ballabio A, Borsani G: **Cloning of the gene encoding a novel integral membrane protein, mucolipidin-1 and identification of the two major founder mutations causing Mucolipidosis type IV.** *Am J Hum Genet* 2000, **67:**1110-1120
9. Sun M, Goldin E, Stahl S, Falardeau JL, Kennedy JC, Acierno JS Jr, Bove C, Kaneski CR, Nagle J, Bromley MC, Colman M, Schiffmann R, Slaugenhaupt SA: **Mucolipidosis type IV is caused by mutations in a gene encoding a novel transient receptor potential channel.** *Hum Mol Genet* 2000, **9:**2471-2478
10. Slaugenhaupt SA, Acierno JS Jr, Helbling LA, Bove C, Goldin E, Bach G, Schiffmann R, Gusella JF: **Mapping of the Mucolipidosis type IV gene to chromosome 19p and defining the founder haplotype.** *Am J Hum Genet* 1999, **65:**773-778
11. Bargal R, Bach G: **Mucolipidosis type IV: abnormal transport of lipids to lysosomes.** *J Inherit Metab Dis* 1997, **20:**625-632
12. Chen CS, Bach G, Pagano RE: **Abnormal transport along the lysosomal pathway in mucolipidosis, type IV disease.** *Proc Natl Acad Sci* 1998, **95:**6373-6378
13. Fares H, Greenwald I: **Regulation of endocytosis by CUP-5, the *Caenorhabditis elegans* mucolipin-1 homolog.** *Nat Genet* 2001, **28:**64-68
14. Kim J, Gordon L, Dehal P, Badri H, Christensen M, Groza M, Ha C, Hammond S, Vargas M, Wehri E, Wagner M, Olsen A, Stubbs L: **Homology-driven assembly of a sequence-ready mouse BAC contig map spanning regions related to the 46-Mb gene-rich euchromatic segments of human chromosome 19.** *Genomics* 2001, **74:**129-141
15. Acierno JS, Kennedy JC, Falardeau JL, Leyne M., Bromley MC, Colman MW, Sun M, Bove C, Ashworth LK, Chadwik LH, Schiripo T, Ma S, Goldin E, Schiffmann R, Slaugenhaupt SA: **A Physical and Transcript Map of the *MCOLN1* Gene Region on Human Chromosome 19p13.3-13.2.** *Genomics* 2001, **73:**203-210
16. Sambrook J, Fritsch EF, Maniatis T: **Molecular cloning: a laboratory manual.** *Cold Spring Harbor, Cold Spring Harbor Laboratory Press* 1989
17. Ballenthin PA, Gardinier MV: **Myelin/oligodendrocyte glycoprotein is alternatively spliced in humans but not mice.** *J Neurosci Res* 1996, **46:**271-281
18. Halleck MS, Schlegel RA, Williamson P: **Re-analysis of ATP11B, a type IV P type ATPase.** *J Biol Chem* 2002