Am. J. Hum. Genet. 72:1359–1369, 2003

Cohen Syndrome Is Caused by Mutations in a Novel Gene, *COH1,* Encoding a Transmembrane Protein with a Presumed Role in Vesicle-Mediated Sorting and Intracellular Protein Transport

Juha Kolehmainen,[1,2] Graeme C. M. Black,[6,7] Anne Saarinen,[1,2] Kate Chandler,[7] Jill Clayton-Smith,[7] Ann-Liz Träskelin,[1,2] Rahat Perveen,[6] Satu Kivitie-Kallio,[3] Reijo Norio,[5] Mette Warburg,[8] Jean-Pierre Fryns,[9] Albert de la Chapelle,[1,10] and Anna-Elina Lehesjoki[1,2,4]

[1]Folkhälsan Institute of Genetics and [2]Department of Medical Genetics, Haartman Institute, and [3]The Hospital for Children and Adolescents, [4]Helsinki University Central Hospital, University of Helsinki, and [5]Department of Medical Genetics, The Family Federation of Finland, Helsinki; [6]Academic Unit of Ophthalmology, University of Manchester, Manchester Royal Eye Hospital, and [7]University Department of Medical Genetics and Regional Genetics Service, St. Mary's Hospital, Manchester, United Kingdom; [8]Centre for Disabled Persons, Glostrup Hospital, Glostrup, Denmark; [9]Center for Human Genetics, University of Leuven, Leuven, Belgium; and [10]Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio

Cohen syndrome is an uncommon autosomal recessive disorder whose diagnosis is based on the clinical picture of nonprogressive psychomotor retardation and microcephaly, characteristic facial features, retinal dystrophy, and intermittent neutropenia. We have refined the critical region on chromosome 8q22 by haplotype analysis, and we report the characterization of a novel gene, *COH1,* that is mutated in patients with Cohen syndrome. The longest transcript (14,093 bp) is widely expressed and is transcribed from 62 exons that span a genomic region of ~864 kb. *COH1* encodes a putative transmembrane protein of 4,022 amino acids, with a complex domain structure. Homology to the *Saccharomyces cerevisiae* VPS13 protein suggests a role for COH1 in vesicle-mediated sorting and transport of proteins within the cell.

Introduction

Cohen syndrome (MIM 216550) (Cohen et al. 1973; Carey and Hall 1978) is one of the rare autosomal recessive disorders that are overrepresented in the Finnish population (Norio 2003). The phenotype in the Finnish patients is highly homogeneous, consisting of nonprogressive mild-to-severe psychomotor retardation, motor clumsiness, microcephaly, characteristic facial features, childhood hypotonia and joint laxity, a progressive retinochoroidal dystrophy, myopia, intermittent isolated neutropenia, and a cheerful disposition (Norio et al. 1984; Kivitie-Kallio and Norio 2001). The characteristic facial features include high-arched or wave-shaped eyelids, a short philtrum, thick hair, and low hairline (fig. 1). The ophthalmological features have been described elsewhere in detail (Kivitie-Kallio et al. 2000); however, in a pleiotropic syndrome like this, it is not possible to determine a minimum number of features required for the diagnosis (Norio et al. 1984; Kivitie-Kallio and Norio 2001).

In non-Finnish patients who are said to have Cohen syndrome, a confusing phenotypic variability prevails. A multitude of different features have been combined with this diagnosis in different patients. Obesity, although frequently mentioned as a characteristic finding, is insignificant. On the other hand, there is no proof of retinochoroidal dystrophy or intermittent neutropenia in reports of several patients. Thus, a distinct clinical—and possibly also genetic—heterogeneity prevails among the patients with Cohen syndrome presented in the literature (Kondo et al. 1990; Kivitie-Kallio and Norio 2001; Chandler and Clayton-Smith 2002).

In studies published elsewhere, we assigned the *COH1* locus for Cohen syndrome to an ~0.6-cM region, close to marker locus *D8S1762* on chromosome 8q22, by use of linkage disequilibrium and haplotype analysis in Finnish families (Tahvanainen et al. 1994; Kolehmainen et al. 1997). In the present study, further haplotype analysis was used to refine the *COH1* critical region, allowing the identification and characterization of a novel gene, *COH1,* that is mutated in patients with Cohen syndrome. It encodes a protein of 4,022 amino acids (aa), of unknown function, whose domain structure and homologies suggest an involvement in vesicle-mediated sorting and intracellular protein transport.

Received December 2, 2002; accepted for publication March 7, 2003; electronically published May 2, 2003.

Address for correspondence and reprints: Dr. Anna-Elina Lehesjoki, Folkhälsan Institute of Genetics, Biomedicum Helsinki, P.O. Box 63 (Haartmaninkatu 8), FIN-00014 University of Helsinki, Helsinki, Finland. E-mail: anna-elina.lehesjoki@helsinki.fi

0002-9297/2003/7206-0002$15.00

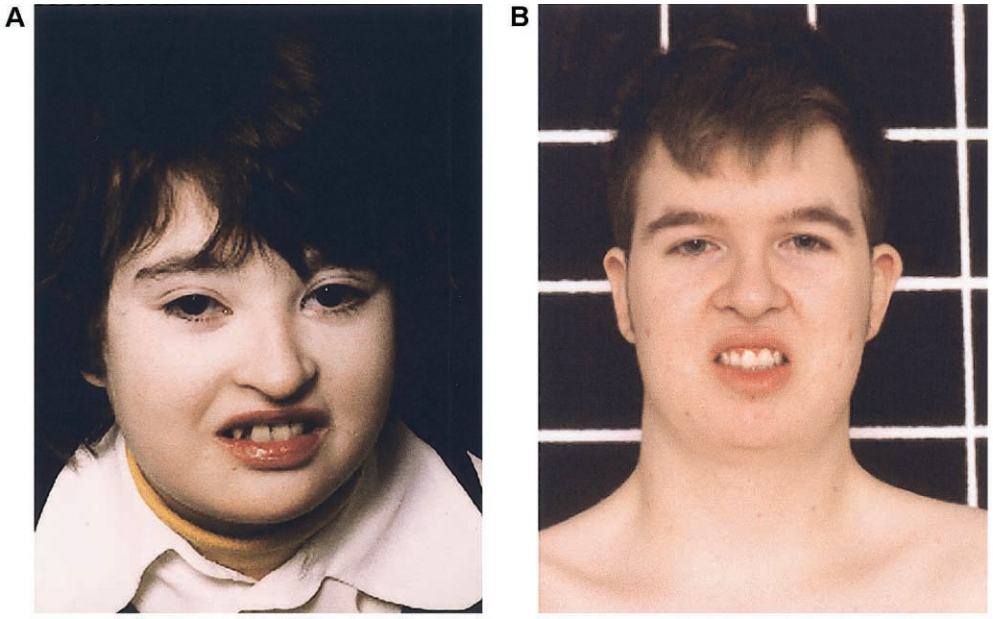

Figure 1 Characteristic facial features in patients with Cohen syndrome. *A*, A 12-year-old girl. *B*, A 21-year-old man. Both patients are Finnish and are homozygous for the c.3348_3349delCT mutation.

Subjects and Methods

Families with Cohen Syndrome and Control Samples

Blood samples were obtained from 27 Finnish patients with Cohen syndrome (patients 1–27) in 20 families and from 43 healthy family members. Part of each blood sample was used for DNA extraction, and part was used to establish a lymphoblastoid cell line for subsequent DNA and RNA extraction. Two patients with Cohen syndrome (patients 28 and 29) from the United Kingdom (Chandler et al. 2003), two (patients 30 and 31) from Belgium (Fryns et al. 1996), and one (patient 32) from Denmark, all of northern European descent, were also included in the study. All patients included in the study had the classical Cohen syndrome phenotype and fulfilled the essential diagnostic criteria for Cohen syndrome, including developmental delay, microcephaly, characteristic facial dysmorphism (fig. 1), retinochoroidal dystrophy, and neutropenia. All but three of the Finnish patients were included in a clinical study reported elsewhere (Kivitie-Kallio and Norio 2001), in which their features are described thoroughly. Control DNA samples from 58 anonymous Finnish blood donors were obtained from the Finnish Red Cross Blood Transfusion Service. DNA samples from 94 grandparents of the CEPH reference families were also used as controls. The study has been reviewed and approved by the ethical review board of the Department of Medical Genetics at the University of Helsinki.

Computational Analyses and Sequence Annotation

Sequence comparisons were performed using BLAST (Altschul et al. 1997) and Align (Southampton Bio-Informatics Data Server; Pearson and Lipman 1988). Prosite (Hofmann et al. 1999) and ProDom (Gouzy et al. 1999) were accessed via the ExPASy Molecular Biology Server. Kyte-Doolittle hydrophobicity profile (Kyte and Doolittle 1982) was analyzed with an accessory application in the BioEdit sequence alignment editor. The secondary structure predictions were performed using SOPMA (Pôle Bio-Informatique Lyonnais; Geourjon and Deleage 1995). PSORT II (ExPASy Molecular Biology Server; Nakai and Kanehisa 1992) was used in the subcellular localization prediction of the protein. Transmembrane topology prediction was performed using TMAP (Persson and Argos 1994). BioEdit accessory application programs CAP3 (Huang and Madan 1999) and Clustal W (Thompson et al. 1994) were used in DNA and protein sequence alignment, respectively.

Identification of the COH1 *Gene*

Continuous partial transcript sequences were assembled *in silico*. Predicted exons and noncontinuous single EST sequences were linked to the *in silico* assembled sequences by RT-PCR on lymphoblastoid cell total RNA. The total assembled gene sequence was verified in overlapping fragments. RNA was isolated using TRIZOL (Life Technologies) and the RNeasy Mini Kit (Qiagen), according to the suppliers' instructions. RT-

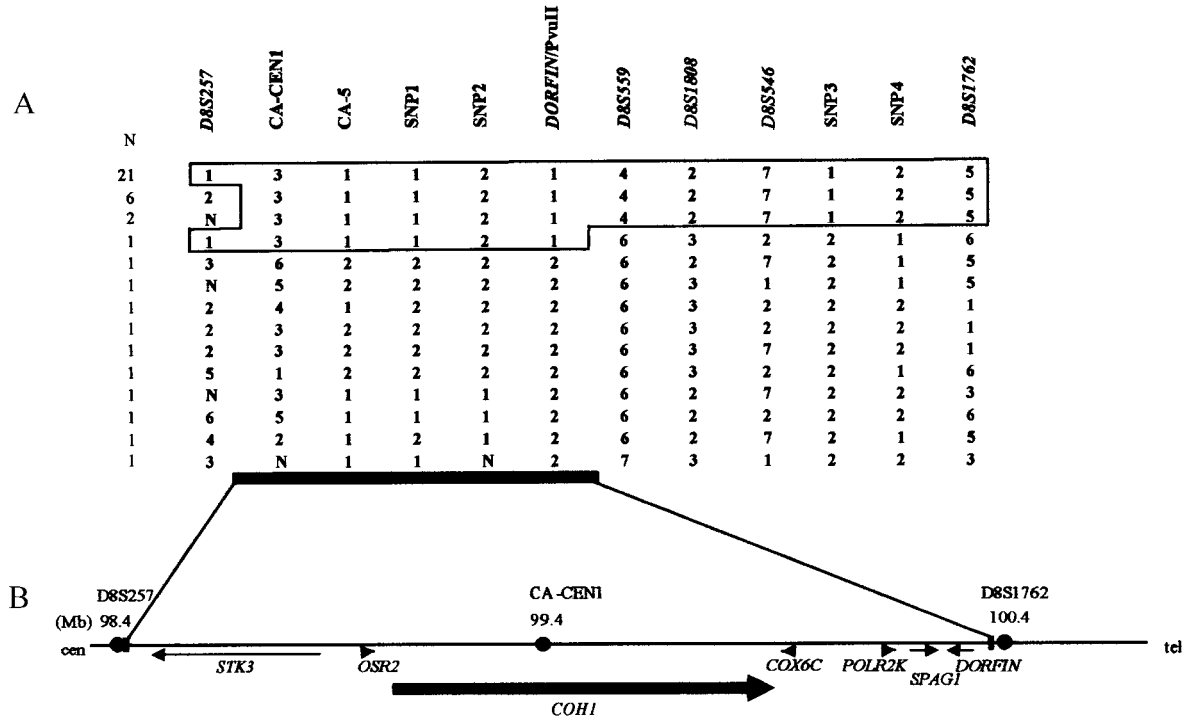

Figure 2 Positional identification of the *COH1* gene. *A*, Haplotypes in 40 *COH1* chromosomes from one patient in each of 20 Finnish families with Cohen syndrome. Twelve markers covering the 2.2-Mb *COH1* critical region were used. Italic type indicates the five microsatellite markers published elsewhere. Marker CA-CEN1 is within intron 33 of *COH1*. Haplotypes associated with the c.3348_3349delCT mutation are boxed. Historical recombinations refine the *COH1* region to between *D8S257* and *D8S559*. Nondefined alleles are marked with "N." *B*, The *COH1* critical interval. Shown are the positions of three polymorphic markers, known genes analyzed for mutations in patients with Cohen syndrome (represented as arrows in the direction of transcription), and the *COH1* transcript (assembled using *in silico* homology searches, exon predictions, and RT-PCR); the direction of transcription is indicated (*arrowhead*).

PCR was performed as described elsewhere (Joensuu et al. 2001).

Mutation Analysis

COH1 was sequenced for mutations in 10 Finnish patients representing different haplotypes and in the five non-Finnish patients. Mutations were identified using either direct sequencing from lymphoblastoid cell RNA after RT-PCR (nine patients) in 27 overlapping fragments or were identified exon by exon (six patients), using direct sequencing or SSCP analysis of genomic DNA (primers available on request). The identified changes were analyzed in all available patients and family members and in control individuals; analysis was performed either by restriction enzyme digestion or by SSCP gel electrophoresis of PCR-amplified genomic DNA.

Northern Blot Analysis

A 489-bp (nt 9488–9976, exons 51–54) fragment of *COH1*, amplified by RT-PCR on lymphoblastoid cell RNA and labeled with [α^{32}P]-dCTP isotope using Rediprime II (Amersham Pharmacia), was used as a probe for northern blot analysis; Nick column (Amersham Pharmacia) purification was performed after labeling. Human multiple-tissue northern (MTN) blots I and II and fetal MTN blots (Clontech) were hybridized in the presence of herring sperm DNA and human placental DNA (Sigma). Hybridization was performed at 65°C after prehybridization in ExpressHyb solution (Clontech). Filters were washed in 1 × SSC and 0.5% SDS solution for 1 h, and x-ray films were developed after exposure at −80°C.

Results

Identification and Characterization of the COH1 *Gene*

Haplotype data in Finnish patients with Cohen syndrome showed that *COH1* is located between marker loci *D8S257* and *D8S559* (fig. 2*A*). Assembly of an ~2.2-Mb sequence contig, including sequencing of a 162-kb BAC clone (Genosys, clone 476J3), allowed the identification of candidate transcripts on the basis of their location (fig. 2*B*). The full sequence of the region between *STK3* and *DORFIN* is available on the November

2002 freeze of the Human Genome Browser (UCSC Genome Bioinformatics). Mutations were defined in Finnish patients within one of these putative transcription units. This transcript was initially identified from homology to mouse BAB26477 (GenBank accession number AK009750). Starting from this sequence, we assembled the full-length *COH1* cDNA (GenBank accession number AY223814) by sequence identification from databases, by exon prediction, and by RT-PCR. Analysis of the full *COH1* transcript suggested a putative ATG translation initiation codon at nucleotide 112 of the *COH1* cDNA sequence although no conserved CAAT, TATA, or Kozak consensus sequences occurred within the available upstream sequence. The transcription initiation site of *COH1* remains to be determined. The longest transcript consists of 14,093 bp and is transcribed from 62 exons spanning a genomic region of ~864 kb (fig. 3*A*). The corresponding ORF encodes a protein of 4,022 aa. Comparison of the *COH1* cDNA sequence with the genomic sequence allowed determination of exon-intron boundaries, all of which carry consensus acceptor and donor splice sites. The translation initiation codon is in exon 2.

There is evidence that the *COH1* gene is alternatively spliced (fig. 3*B*). Overlapping EST sequences suggest alternative splicing of exons 8, 17, 28, and 31 (e.g., AI377522 for alternative exon 8b, BM042444 for alternative exon 17b, BE048857 for alternative exon 28b, AI809885 for the presence of both exon 28 and exon 28b, and BF767258 for the lack of exon 31). All but the variant lacking the 204-bp exon 31 were confirmed by RT-PCR (data not shown). The use of alternative exons 8b, 17b, or 28 together with 28b would be predicted to result in truncated forms of COH1, whereas exons 28, 28b, and 31 are spliced in frame (fig. 3*B*). Northern blot analysis, using a cDNA probe (exons 51–54), showed *COH1* to be widely expressed (fig. 4) although there was apparent differential expression of different transcripts. In fetal brain, lung, liver, and kidney, two transcripts of ~2 and ~5 kb were identified.

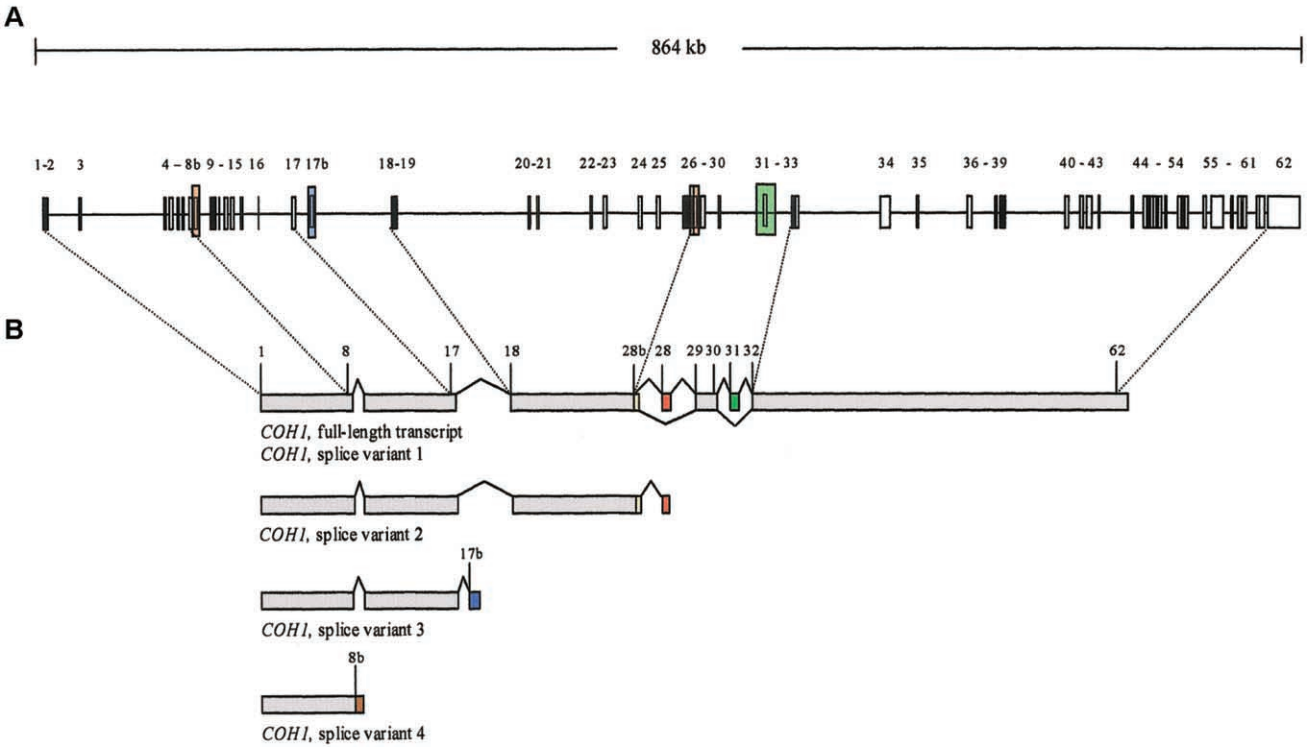

Figure 3 Genomic structure of *COH1* and schematic representation of *COH1* alternative splicing. *A*, Genomic structure of *COH1*. The *COH1* transcript, comprising 14,093 nt, covers an 864-kb genomic region and is transcribed from 62 exons. The alternatively spliced exons present in ESTs are shown as colored boxes. The exons are shown in scale relative to each other. The lengths of the intronic sequences are in scale relative to each other but not relative to the exons. *B*, Schematic representation of alternative splicing of *COH1*. The longest predicted transcript (GenBank accession number AY223814) uses exons 8, 17, 28, and 31 and has an ORF of 12,066 bp. Use of exon 28b, instead of 28, results in a transcript (GenBank accession number AY223815) that is 75 bp shorter. Additional variation is predicted from in-frame splicing of the 204-bp exon 31 present in ESTs but not experimentally verified. The presence of exons 28 and 28b predicts a shorter transcript, with an ORF of 4,281 bp (GenBank accession number AY223816). Alternative exons 17b and 8b result in ORFs of 2,589 bp (Genbank accession number AY223817) and 1,236 bp (GenBank accession number AY223818), respectively.

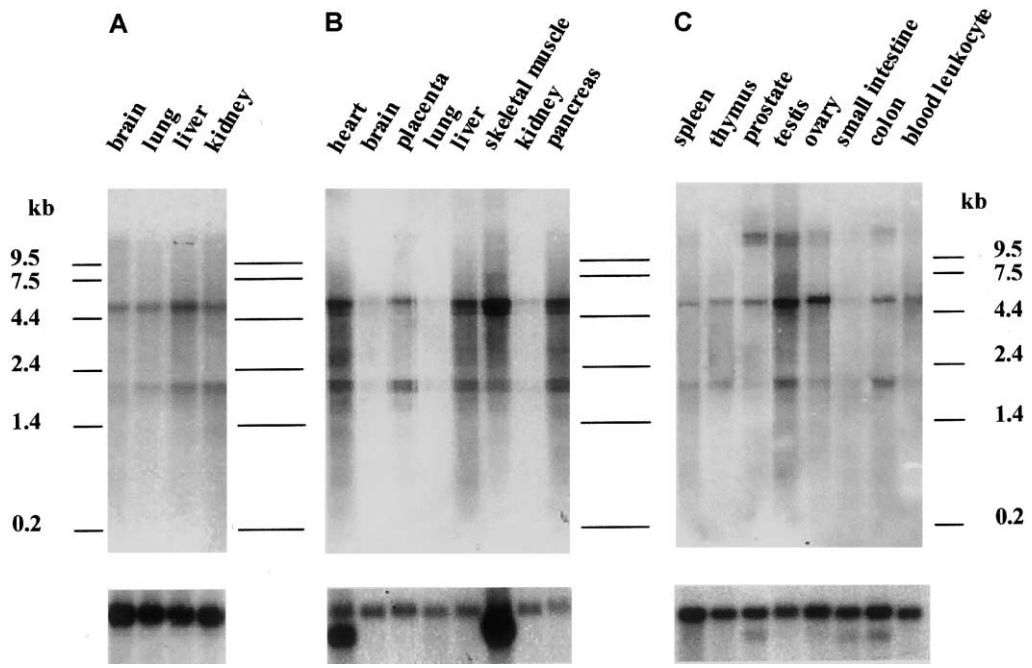

Figure 4 Expression pattern of the *COH1* gene. *COH1* is widely expressed. A 489-bp probe from the coding region of *COH1* was used in human MTN blot analysis (fetal MTN [*A*], MTN I [*B*], and MTN II [*C*]; Clontech). Two isoforms (~2 and ~5 kb) were present in all tissues. Expression of the longest ~12–14-kb transcript was detectable in prostate, testis, ovary, and colon. A β-actin probe (*bottom*) was used as a control for RNA loading.

These transcripts were also seen in all adult tissues analyzed. A larger transcript (~12–14 kb) was expressed in prostate, testis, ovary, and colon in the adult. Expression of *COH1* was very low in adult brain tissue.

Identification of COH1 *Mutations*

We analyzed 27 Finnish patients with Cohen syndrome, from 20 families, as well as five non-Finnish patients. The results are given in table 1 and depicted in figure 5. Haplotype analysis of the Finnish patients suggested the existence of several mutations. One haplotype, present in 30/40 (75%) of disease chromosomes (fig. 2*A*, boxed haplotypes), was always found to be associated with a 2-bp (CT) deletion affecting codons 1116 and 1117 (c.3348_3349delCT; fig. 5*A*) that leads to protein truncation at codon 1124. The mutation cosegregated with the disease phenotype in families with Cohen syndrome. Patients in 11 families (patients 1–15) were homozygous for this mutation. It was also found in heterozygous form in eight families (patients 16–26), in which the patients' other *COH1* chromosome carried a different haplotype. In only one of these families (patient 16) was a further putative mutation identified, a missense substitution (c.6578T→G at codon 2193; fig. 5*B*) of a neutral leucine to a basic arginine residue. Neither mutation was found in 58 Finnish or 94 CEPH control samples. Only one Finnish family (patient 27) did not carry the most common haplotype and is likely to be homozygous for a further, as yet unidentified *COH1* mutation.

The seven mutations identified in five non-Finnish patients with Cohen syndrome (table 1; fig. 5*C*–5*I*) are predicted to result in premature protein truncation. One Belgian patient (patient 30) was homozygous for an R2351X mutation in exon 39 (fig. 5*C*). Two patients, one British (patient 28) and one Belgian (patient 31) were compound heterozygotes (table 1; fig. 5*D*–5*G*). In addition, two other protein-truncating mutations were found in heterozygous form (patients 29 and 32; figs. 5*H* and 5*I*). None of the non-Finnish *COH1* mutations identified was found in 94 CEPH controls.

Protein Homologies and Predicted Domains of the COH1 Protein

Analysis of aa sequence of the predicted COH1 protein (fig. 6), using ProDom, revealed high similarity (94%) over aa 2–317 to domain family PD149545 and revealed moderate similarity (49%; aa 3711–3824) to domain family PD025730. Both are homologous to *Saccharomyces cerevisiae* vacuolar sorting-associated protein VPS13. Moreover, a Prosite scan revealed 100% pattern homologies (aa 3637–3665) to the binding-protein–dependent transport systems inner membrane component signature domain (Higgins et al. 1990) and to

Table 1

Summary of *COH1* Gene Mutations in 32 Patients

Mutation		Exon	Patient No.[a]	No. of Homozygous/ Heterozygous Patients
DNA Change	Protein Change			
c.3348_3349delCT	C1117F…I1124X	23	1–26	15/11
c.6578T→G	L2193R	37	16	0/1
c.4572_4573insA	E1525R…E1569X	29	28	0/1
c.EX30del[b]	A1570G…A1573X	30	28	0/1
c.8472G→A	W2824X	46	29	0/1
c.7051C→T	R2351X	39	30	1/0
c.5426_5427dupAG	Q1810S…K1830X	34	31	0/1
c.2193C→T	R692X	15	31	0/1
c.6420_6421delGA	Q2140H…L2167X	36	32	0/1

[a] No mutation had been identified in patient 27.
[b] The genomic basis of c.EX30del remains undefined, because the splice sites and 360 bp of intronic sequences flanking exon 30 were normal.

the zinc-binding region signature of the neutral zinc metallopeptidases (Jongeneel et al. 1989). A possible vacuolar targeting motif (Stack et al. 1995; aa 1518–1521) was found with the PSORT II program, which also predicted COH1 to have endoplasmic reticulum (ER) membrane topology. In addition, TMAP and Kyte-Doolittle hydrophobicity analysis predicted 10 transmembrane domains for COH1 (fig. 6). COH1 contains a potential ER retention signal (Teasdale and Jackson 1996) at the C-terminus, with a lysine residue at aa −4. Two peroxisomal matrix protein targeting signal 2 (PTS2) consensus sequences (McNew and Goodman 1996) were found, one close to the N-terminus (aa 263–271) and the other close to the C-terminus (aa 3553–3561).

Protein BLAST alignment of the COH1 sequence showed cross-species homology to several proteins of unknown function. In particular, the N- and C-terminal sequences (aa 1–85 and 3,639–3,826) were conserved and shared homology with *S. cerevisiae* VPS13 (Q07878) protein, as well as the *CHAC* gene underlying choreoacanthocytosis (MIM 200150) (Rampoldi et al. 2001; Ueno et al. 2001). This indicates a possible involvement in control of trans-Golgi network protein sorting.

Discussion

We have identified a gene, *COH1,* spanning a genomic region of 864 kb that is mutated in Cohen syndrome and encodes a novel protein of unknown function. We had previously mapped the *COH1* locus close to marker *D8S1762* on chromosome 8q22, by linkage and linkage disequilibrium analysis (Tahvanainen et al. 1994; Kolehmainen et al. 1997). A historical recombination in a single disease chromosome in a patient with newly diagnosed Cohen syndrome (fig. 2*A*) excluded the region distal to the *DORFIN* gene as the site of the causative mutation, allowing us to focus the search for *COH1* in the region between loci *D8S257* and *D8S559*. We have now identified a total of nine putative mutations within the *COH1* gene, eight of which are predicted to result in premature protein truncation. Within the Finnish cohort, the mutations identified confirmed our haplotype analyses and suggested that a number of mutations are still to be identified. The mutation-detection techniques employed to date are not 100% sensitive, and, in particular, large intragenic genomic deletions in compound heterozygous patients would not have been identified by the analyses of genomic DNA. This remains an avenue for future investigation. The nonconserved missense substitution observed in one Finnish patient (patient 16; table 1), was not present in 304 control chromosomes. However, it remains possible that this is a rare polymorphism rather than a pathogenic mutation. All patients screened had the classical Cohen syndrome phenotype, as defined in Finnish families (Kivitie-Kallio and Norio 2001). All 31 patients with mutations in *COH1* (table 1) show a striking clinical similarity, with clinical homogeneity among Finnish and non-Finnish patients with Cohen syndrome. Their phenotype includes microcephaly (100%), developmental delay (100%), retinal dystrophy (100%), neutropenia (100%), joint laxity (97%; 31/32), and the characteristic dysmorphic appearance of Cohen syndrome (100%). Mutational studies will now be required to show the extent to which the clinical heterogeneity among other reported patients with Cohen syndrome is explained by allelic or locus heterogeneity.

Examination of homologous EST sequences shows that the *COH1* gene has a complicated pattern of alternative splicing which potentially leads to the use of four different termination codons (fig. 3*B*) and to three additional in-frame, alternatively spliced forms. In addition, computational analysis indicates a complex, multidomain structure for the COH1 protein, implying that the various COH1 isoforms may have different

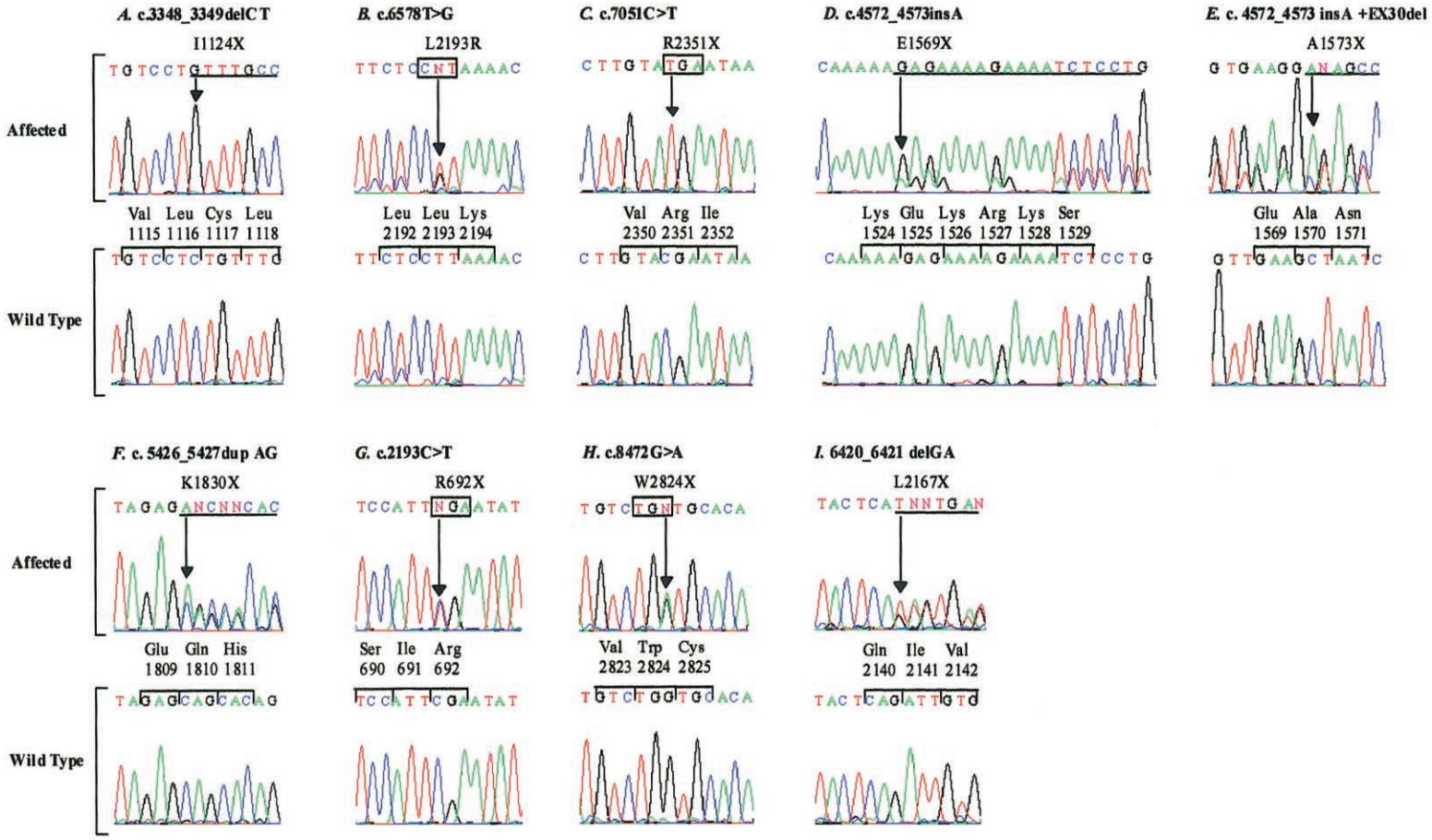

Figure 5 Mutations in the *COH1* gene. *A–I*, sequencing chromatograms showing the nine identified mutations from seven patients with Cohen syndrome and controls. Two mutations (*A* and *C*) are homozygous, whereas the rest are in heterozygous form. Nucleotide changes (*arrows*) and codons with a single aa change (*box*) are shown. Out-of-frame sequences are underlined.

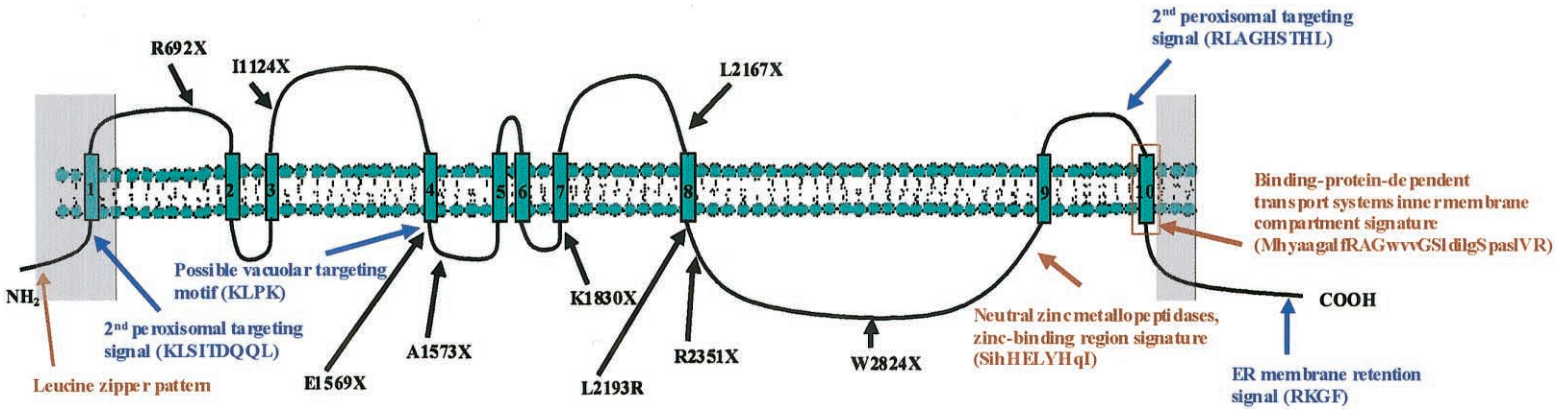

Figure 6 Predicted membrane topology and domains of COH1. The predicted transmembrane domains are numbered 1–10. The locations of the mutations identified in patients with Cohen syndrome are shown (*blackened arrows*). The positions of the predicted functional motifs (*blue arrows*) and vacuolar sorting–associated VPS13 domains (*gray boxes*) are indicated. Conserved patterns observed in Prosite database searches are also indicated (*brown arrows*).

functions within the cell. The functional relevance of the predicted domains needs to be tested experimentally. However, on the basis of the domains present and the identified homologies of the COH1 protein, some functional predictions can be made. The conserved N- and C-terminal domains share homology to the *S. cerevisiae* VPS13 protein. The VPS13 protein in *S. cerevisiae* is known to be involved in trafficking of membrane proteins between the trans-Golgi network and the prevacuolar compartment (Stack et al. 1995; Redding et al. 1996; Brickner and Fuller 1997). This suggests a role in intracellular vesicle-mediated sorting and transport of proteins. The presence of subcellular targeting motifs may suggest a role in protein sorting at various subcellular locations. COH1 contains two peroxisomal matrix protein targeting signal 2 (PTS2) consensus sequences, which, contrary to known peroxisomal proteins with a PTS2 signal (Erdmann 1994; Braverman et al. 1997; Purdue et al. 1997), are located in atypical positions. To our knowledge, the coexistence of ER retention and matrix PTS2 signals has not been previously described, whereas several peroxisomal membrane proteins have been shown to have overlapping ER-targeting and peroxisomal membrane–targeting signals, suggesting direct involvement of ER in regulating the sorting of proteins to peroxisomes (Mullen et al. 1999; Mullen and Trelease 2000).

High homology of the N- and C-terminal sequences of COH1 to the relatively large VPS13-like proteins in eight species (*S. cerevisiae*, *Schizosaccharomyces pombe, Caenorhabditis elegans, Drosophila melanogaster, Arabidopsis thaliana, Anopheles gambiae, Macaca fascicularis,* and *Dictyostelium discoideum*) indicate that the corresponding genes are evolutionarily related to *COH1*. A human gene with homologies to *COH1, CHAC* (which, when mutated, underlies choreoacanthocytosis [Rampoldi et al. 2001; Ueno et al. 2001]) has been described elsewhere to encode a protein, chorein, with N- and C-terminal VPS13 domains. It is perhaps of interest that both Cohen syndrome and choreoacanthocytosis involve similar organ systems (psychomotor deficiency and a hematological abnormality); this suggests the intriguing possibility that the two diseases may share a common pathogenic pathway. The motifs predicted from the VPS13-like proteins reveal a variable group of proteins, some of which display predicted peroxisomal, ER, and vacuolar targeting signals like COH1 (e.g., *D. melanogaster* protein CG32113-PA), whereas some lack these targeting signals (e.g., human chorein and yeast VPS13).

Identification of the gene underlying Cohen syndrome will facilitate the early diagnosis of the disorder and allow the definition of the phenotypic range of the condition. The COH1 protein identified is novel and of unknown function, and its possible involvement in vesicle-mediated sorting and intracellular protein transport may indicate a family of proteins that includes candidates for other mental retardation syndromes. Future work must focus on understanding the function of COH1 and on identifying the pathogenic mechanism underlying Cohen syndrome.

Acknowledgments

We thank the patients with Cohen syndrome and their families, for collaboration; S. Lindh, for collecting the blood samples; and P. Sistonen, for the Finnish Red Cross Blood Transfusion Service control samples. Sequencing and help with sequence assembly were provided by R. B. Chadwick, B. Yuan, C. Johnson, and S. T. Richards at the Ohio State University. This study was supported by the Ulla Hjelt Foundation, the Folkhälsan Institute of Genetics, and the Finnish Medical Society Duodecim and was performed in the Centre of Excellence in Disease Genetics of the Academy of Finland (project 44870, Finnish Centre of Excellence Programme 2000–2005). The work was also supported by National Cancer Institute grant P30 CA 16058. G.C.M.B. is a Wellcome Trust Senior Clinical Research Fellow.

Electronic-Database Information

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

BioEdit, http://www.mbio.ncsu.edu/BioEdit/bioedit.html (for CAP3 and Clustal W)

BLAST, http://www3.ncbi.nlm.nih.gov/BLAST/

Celera, http://www.celera.com/

ExPASy Molecular Biology Server, http://www.expasy.ch/ (for PSORT II, PROSITE, and ProDom)

GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for *DORFIN* [*Homo sapiens*] mRNA [accession number AB029316]; BAB26477 [*Mus musculus*] mRNA [accession number AK009750]; *COH1*[*H. sapiens*] mRNA [accession number AY223814]; vps13 [*S. cerevisiae*] protein [accession number Q07878]; *chorein* [*H. sapiens*] protein [accession number Q96RL7]; CG32113-PA [*D. melanogaster*] protein [accession number AAF49887]; *COH1,* splice variant 1 [*H. sapiens*] [accession number AY223815] mRNA; *COH1,* splice variant 2 [*H. sapiens*] [accession number AY223816] mRNA; *COH1,* splice variant 3 [*H. sapiens*] [accession number AY223817] mRNA; and *COH1*, splice variant 4 [*H. sapiens*] [accession number AY223818] mRNA)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for Cohen syndrome and choreoacanthocytosis)

Pôle Bio-Informatique Lyonnais, http://npsa-pbil.ibcp.fr/ (for SOPMA)

Southampton BioInformatics Data Server, http://molbiol.soton.ac.uk/ (for Align)

TMAP Multiple Sequence Alignment Form, http://www.mbb.ki.se/tmap/

UCSC Genome Bioinformatics, http://genome.cse.ucsc.edu/ (for Human Genome Browser)

UK Human Genome Mapping Project Resource Centre, http://www.hgmp.mrc.ac.uk/ (for NIX)

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Braverman N, Steel G, Obie C, Moser A, Moser H, Gould SJ, Valle D (1997) Human PEX7 encodes the peroxisomal PTS2 receptor and is responsible rhizomelic chondrodysplasia punctata. Nat Genet 15:369–376

Brickner JH, Fuller RS (1997) SOI1 encodes a novel, conserved protein that promotes TGN-endosomal cycling of Kex2p and other membrane proteins by modulating the function of two TGN localization signals. J Cell Biol 139:23–36

Carey JC, Hall BD (1978) Confirmation of the Cohen syndrome. J Pediatr 93:239–244

Chandler KE, Clayton-Smith J (2002) Does a Jewish type of Cohen syndrome truly exist? Am J Med Genet 111:453–454

Chandler KE, Kidd A, Al-Gazali L, Kolehmainen J, Lehesjoki A-E, Black GCM, Clayton-Smith J (2003) Diagnostic criteria, clinical characteristics and natural history of Cohen syndrome. J Med Genet 40:233–241

Cohen MM Jr, Hall BD, Smith DW, Graham CB, Lampert KJ (1973) A new syndrome with hypotonia, obesity, mental deficiency, and facial, oral, ocular and limb anomalies. J Pediatr 83:280–284

Erdmann R (1994) The peroxisomal targeting signal of 3-oxoacyl-CoA thiolase from *Saccharomyces cerevisiae*. Yeast 10:935–944

Fryns JP, Legius E, Devriendt K, Meire F, Standaert L, Baten E, Van den Berghe H (1996) Cohen syndrome: the clinical symptoms and stigmata at a young age. Clin Genet 49:237–241

Geourjon C, Deleage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 11:681–684

Gouzy J, Corpet F, Kahn D (1999) Whole genome protein domain analysis using a new method for domain clustering. Comput Chem 23:333–340

Higgins CF, Hyde SC, Mimmack MM, Gileadi U, Gill DR, Gallagher MP (1990) Binding protein-dependent transport systems. J Bioenerg Biomembr 22:571–592

Hofmann K, Bucher P, Falquet L, Bairoch A (1999) The PROSITE database, its status in 1999. Nucleic Acids Res 27: 215–219

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9:868–877

Joensuu T, Hämäläinen R, Yuan B, Johnson C, Tegelberg S, Gasparini P, Zelante L, Pirvola U, Pakarinen L, Lehesjoki AE, de la Chapelle A, Sankila EM (2001) Mutations in a novel gene with transmembrane domains underlie Usher syndrome type 3. Am J Hum Genet 69:673–684

Jongeneel CV, Bouvier J, Bairoch A (1989) A unique signature identifies a family of zinc dependent metallopeptidases. FEBS Lett 242:211–214

Kivitie-Kallio S, Norio R (2001) Cohen syndrome: essential features, natural history, and heterogeneity. Am J Med Genet 102:125–135

Kivitie-Kallio S, Summanen P, Raitta C, Norio R (2000) Ophthalmologic findings in Cohen syndrome: a long-term follow-up. Ophthalmology 107:1737–1745

Kolehmainen J, Norio R, Kivitie-Kallio S, Tahvanainen E, de la Chapelle A, Lehesjoki A-E (1997) Refined mapping of the Cohen syndrome gene by linkage disequilibrium. Eur J Hum Genet 5:206–213

Kondo I, Nagataki S, Miyagi N (1990) The Cohen syndrome: does mottled retina separate a Finnish and a Jewish type? Am J Med Genet 37:109–113

Kyte J, Doolittle RF (1982) A simple method for displaying the hydrophobic character of a protein. J Mol Biol 157:105–132

McNew JA, Goodman JM (1996) The targeting and assembly of peroxisomal proteins: some old rules do not apply. Trends Biochem Sci 21:54–58

Mullen RT, Lisenbee CS, Miernyk JA, Trelease RN (1999) Peroxisomal membrane ascorbate peroxidase is sorted to a membranous network that resembles a subdomain of the endoplasmic reticulum. Plant Cell 11:2167–2185

Mullen RT, Trelease RN (2000) The sorting signals for peroxisomal membrane-bound ascorbate peroxidase are within its C-terminal tail. J Biol Chem 275:16337-16344

Nakai K, Kanehisa M (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics 14: 897–911

Norio R (2003) The Finnish disease heritage. I. Characteristics, causes, background. Hum Genet112:441–456

Norio R, Raitta C, Lindahl E (1984) Further delineation of the Cohen syndrome: report on chorioretinal dystrophy, leukopenia and consanguinity. Clin Genet 25:1–14

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85:2444–2448

Persson B, Argos P (1994) Prediction of transmembrane segments in proteins utilising multiple sequence alignments. J Mol Biol 237:182–192

Purdue PE, Zhang JW, Skoneczny M, Lazarow PB (1997) Rhizomelic chondrodysplasia punctata is caused by deficiency of human PEX7, a homologue of the yeast PTS2 receptor. Nat Genet 15:381–384

Rampoldi L, Dobson-Stone C, Rubio JP, Danek A, Chalmers RM, Wood NW, Verellen C, Ferrer X, Malandrini A, Fabrizi GM, Brown R, Vance J, Pericak-Vance M, Rudolf G, Carre S, Alonso E, Manfredi M, Nemeth AH, Monaco AP (2001) A conserved sorting-associated protein is mutant in chorea-acanthocytosis. Nat Genet 28:119–120

Redding K, Brickner JH, Marschall LG, Nichols JW, Fuller RS (1996) Allele-specific suppression of a defective trans-Golgi network (TGN) localization signal in Kex2p identifies three genes involved in localization of TGN transmembrane proteins. Mol Cell Biol 16:6208–6217

Stack JH, Horazdovsky B, Emr SD (1995) Receptor-mediated protein sorting to the vacuole in yeast: roles for a protein kinase, a lipid kinase and GTP-binding proteins. Annu Rev Cell Dev Biol 11:1–33

Tahvanainen E, Norio R, Karila E, Ranta S, Weissenbach J,

Sistonen P, de la Chapelle A (1994) Cohen syndrome gene assigned to the long arm of chromosome 8 by linkage analysis. Nat Genet 7:201–204

Teasdale RD, Jackson MR (1996) Signal-mediated sorting of membrane proteins between the endoplasmic reticulum and the golgi apparatus. Annu Rev Cell Dev Biol 12:27–54

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680

Ueno S, Maruki Y, Nakamura M, Tomemori Y, Kamae K, Tanabe H, Yamashita Y, Matsuda S, Kaneko S, Sano A (2001) The gene encoding a newly discovered protein, chorein, is mutated in chorea-acanthocytosis. Nat Genet 28: 121–122