

Identification of the gene causing mucopolipidosis type IV

Ruth Bargal^{1*}, Nili Avidan^{2*}, Edna Ben-Asher², Zvia Olender², Marcia Zeigler¹, Ayala Frumkin¹, Annick Raas-Rothschild¹, Gustavo Glusman², Doron Lancet² & Gideon Bach¹

*These authors contributed equally to this work.

Mucopolipidosis type IV (MLIV) is an autosomal recessive, neurodegenerative, lysosomal storage disorder¹ characterized by psychomotor retardation and ophthalmological abnormalities including corneal opacities, retinal degeneration and strabismus. Most patients reach a maximal developmental level of 12–15 months². The disease was classified as a mucopolipidosis following observations by electron microscopy indicating the lysosomal storage of lipids together with water-soluble, granulated substances^{1,3–6}. Over 80% of the MLIV patients diagnosed are Ashkenazi Jews, including severely affected and mildly affected patients^{3,4}. The gene causing MLIV was previously mapped to human chromosome 19p13.2–13.3 in a region of approximately 1 cM (ref. 7). Haplotype analysis in the MLIV gene region of over 70 MLIV Ashkenazi chromosomes indicated the existence of two founder chromosomes among 95% of the Ashkenazi MLIV families: a major haplotype in 72% and a minor haplotype in 23% of the MLIV chromosomes (ref. 7, and G.B., unpublished data). The remaining 5% are distinct haplotypes found only in single patients. The basic metabolic defect causing the lysosomal storage in MLIV has not yet been identified. Thus, positional cloning was an alternative to identify the MLIV gene. We report here the identification of a new gene in this human chromosomal region in which MLIV-specific mutations were identified.

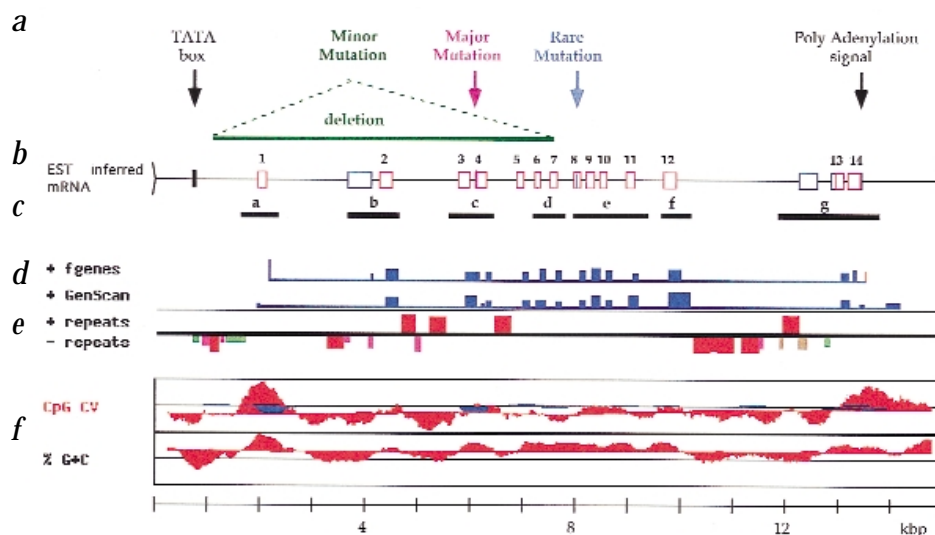
We identified 5 genomic clones using the polymorphic markers covering the MLIV locus interval, between *D19S406* and *D19S76* (ref. 7, and G.B., unpublished data), and the chromosome 19 map (Lawrence Livermore National Laboratory). One of these clones contained the three polymorphic markers, *D19S395*, *D19S901* and *D19S873*, previously reported to be in linkage disequilibrium with the MLIV locus (ref. 7, and G.B., unpublished data). We used the GESTALT Workbench⁸ to analyse the unfinished genomic sequence

of this clone, which included 70 unordered contigs. This indicated the presence of three genes within this genomic clone. Homology search of these contigs identified two previously known genes, *NTE* (encoding neuropathy target esterase) and *KIAA0521* (encoding a putative Rho-guanine nucleotide exchange factor). We analysed thoroughly all coding regions of these genes, 33 and 20 exons, respectively, in 6 MLIV patients of different haplotypes and did not identify any mutations causing MLIV (data not shown).

The third gene was predicted by the GenScan⁹ and fgenes¹⁰ programs within GESTALT (Fig. 1d), and identified as the MLIV gene (*MCOLN1*). A BLAST (ref. 11) search of this genomic region showed identity to an array of ESTs with sufficient overlaps to allow the construction of a putative mRNA of 2,004 nt. Because only overlapping ESTs were used for this assembly, we ignored some irrelevant ESTs in this region. Alignment between this mRNA and the genomic sequence showed that the gene consists of 14 exons spanning approximately 14 kb of genomic DNA (Fig. 1b). This structure was highly correlated with the exon prediction of fgenes (Fig. 1d). We analysed the *MCOLN1* cDNA by PCR amplification between exons 1 and 13 and exons 2 and 14, followed by DNA sequencing that confirmed the gene structure along these exons (for exon-intron boundaries, see Table 1, http://genetics.nature.com/supplementary_info/). Further analysis of the genomic sequence indicated a high GC content, as well as a CpG island upstream from the first exon, supporting the presence of a promoter (Fig. 1f).

The gene was analysed for MLIV-causing mutations (Fig. 1c, fragments a–g). We identified three mutations in MLIV patients (Fig. 1a) in correlation with their haplotypes. The first, an A→G transition in the acceptor splice site of the third intron (IVS3-1A→G), was associated with the major Ashkenazi haplotype⁷ (Fig.

Fig. 1 MLIV gene structure. **a**, Genetic landmarks. **b**, Exons (red boxes) inferred by EST homology searches. EST sequences not included in the putative mRNA are indicated by blue boxes. **c**, Genomic segments amplified by PCR for mutation analysis. **d**, Gene prediction results (fgenes and GenScan). Predicted exons are displayed in blue, with box height indicating exon quality (the scaling is arbitrary but consistent for each prediction program). Complete gene structures are underlined in blue; poly(A) signal is indicated in red. **e**, Repetitive sequences. *Alu* repeats are denoted in red; MIRs, in purple; LINES, in green; other interspersed repeats, in brown. **f**, Compositional analyses. CpG contrast values and %G+C are displayed as deviations from the regional average (GESTALT analysis⁹).



¹Department of Human Genetics, Hadassah Hebrew University Hospital, Jerusalem. ²The Crown Human Genome Center, Department of Molecular Genetics, The Weizmann Institute of Science, Rehovot, Israel. Correspondence should be addressed to G.B. (e-mail: Bach@hadassah.org.il).

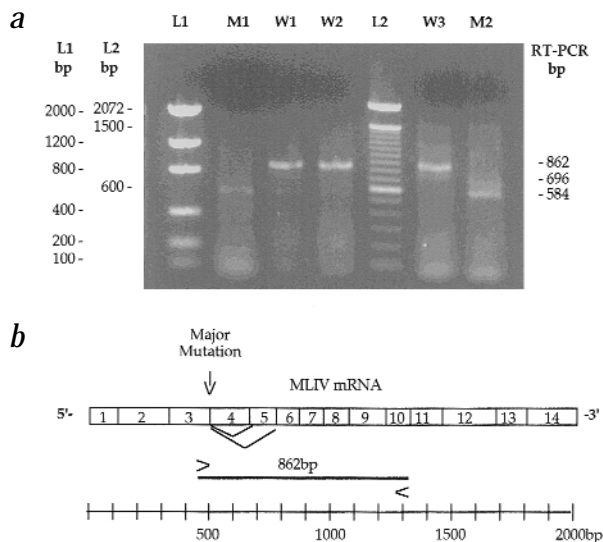


Fig. 2 Splice changes of the major MLIV haplotype. **a**, RT-PCR products of the major haplotype (M) and unaffected individuals (W) are compared. The expected sizes of the RT-PCR products of the unaffected individuals, as well as the respective products of two alternative exon skipping events, are indicated on the right. L, DNA markers. **b**, The exon structure of *MCOLN1* with an indication of the PCR amplification boundaries and skipped exons.

2b). To assess aberrant splicing, we carried out RT-PCR analysis between exons 3 and 10 (Fig. 2b). Shorter products of 696 nt, skipping exon 4 (minor band), and 584 nt, skipping exons 4+5 (major band and the preferred product in this mutation), were observed (Fig. 2a). This mutation creates a *KpnI* site in genomic DNA which identifies heterozygotes and homozygotes (Fig. 3). The mutation was not observed in other MLIV haplotypes. We analysed 60 Ashkenazi normal controls and identified 1 person to be heterozygous for this mutation, as expected from the estimated frequency of heterozygotes (1/50) for the disease in this population¹².

The second MLIV-causing mutation, which is associated with the minor Ashkenazi haplotype, was identified as a deletion of 6,450 bp, resulting in the absence of fragments a–d from the homozygote (Figs 1c and 4). Sequence analysis indicated that the deletion spans a region from 928 bp upstream from the first exon of *MCOLN1* to the 31 bp of exon 7 (del(EX1-EX7); Fig. 1a). These patients do not lack any of the exons of the two genes, *NTE* and *KIAA0521*, or any of the polymorphic markers analysed. It should be noted that for the genomic *KpnI* restriction analysis, compound heterozygotes for the two mutations appear to be homozygous for the splice mutation due to the deletion of this region in the second mutation (Fig. 3). The clinical manifestations of homozygotes for either of the two mutations, or compound heterozygotes, showed similar severity (G.B., unpublished data). We studied 21 Ashkenazi MLIV patients for the 2 muta-

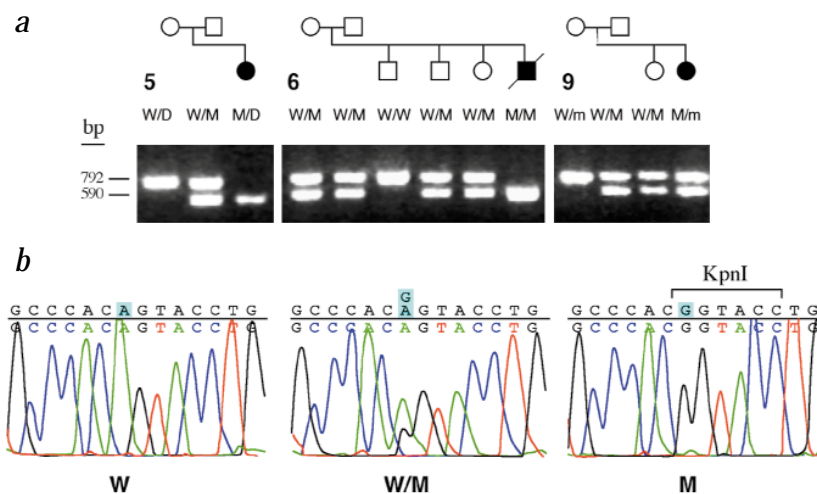
tions and found 12 homozygotes for the splice mutation, 1 homozygote for the del(EX1-EX7) mutation and 6 compound heterozygotes. Two patients were compound heterozygotes for one of these mutations and a yet unidentified second allele.

We identified a nonsense mutation (c.1048C>T, pR321X) causing premature termination in exon 8 in an Arab-Druze MLIV patient. The patient is homozygous for this mutation, as his parents are first cousins bearing the same unique haplotype.

MCOLN1 has an ORF of 1,740 bp (r.nt 85–1825). Northern-blot analysis revealed a single mRNA species of 2 kb, expressed in variable amounts in all tissues examined, including brain, skeletal muscle, colon, thymus, liver, lung and leukocytes (data not shown). The sequence of the entire cDNA, that is between exons 1 and 14, has been verified. The gene encodes a protein of 580 amino acids, which we have named mucolipin-1 (Fig. 5). Analysis of this protein using Kyte Doolittle hydrophathy plot¹³ and PredictProtein¹⁴ indicated one transmembrane helix in the amino-terminal region and at least five transmembrane domains in the carboxy-terminal half of the protein (aa 350–580; see Fig. 7, http://genetics.nature.com/supplementary_info/). Sequence comparison of the human mucolipin-1 protein revealed similarity to an entire family of new human proteins as well as to potential orthologues in *Drosophila melanogaster* and *Caenorhabditis elegans* (see Fig. 7, http://genetics.nature.com/supplementary_info/). In addition, mucolipin-1 showed similarity to members of the polycystin protein family¹⁵ (data not shown). Analysis with the PRSS software¹⁶ showed that the above sequence similarities are significant and do not stem merely from compositional similarity of the transmembrane regions. Polycystins were predicted to be cation channels¹⁷. Thus, it is possible that the mucolipins encode new ion channel proteins, a conclusion strengthened by the finding of a region with multiple putative transmembrane domains.

In addition, we performed a PROSITE analysis^{18,19} and found that the first half of the protein (aa 1–270) contains a consensus motif of lipases with serine active site at amino acids 104–113 (Fig. 6), as well as bipartite nuclear localization signal (aa 41–60; see Fig. 7, http://genetics.nature.com/supplementary_info/). Therefore, this protein may have a lipid-related enzymatic function, in potential correlation with the cell biology of the MLIV disease.

Fig. 3 *KpnI* restriction analysis of the splice mutation in MLIV families. **a**, In family 5, the patient is a compound heterozygote for the splice mutation (M) and the deletion (D); his mother carries the deletion. In family 6, the patient is homozygous for the splice mutation and his parents are heterozygotes. In family 9, the patient is a compound heterozygote for the splice mutation and an unknown mutation (m). The mother carries the unknown mutation. The unaffected sister carries the normal chromosome from her mother. The PCR fragment for the *KpnI* analysis is 792 bp, cutting with *KpnI* results in two bands, one of 590 bp and another of 202 bp (not shown). **b**, The chromatograms of a normal control (W), a heterozygote (W/M) and the mutation (M) creating a *KpnI* site.



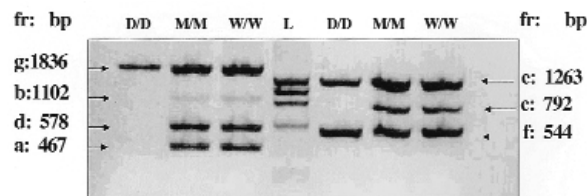


Fig. 4 Presentation of the deletion in the minor MLIV mutation. PCR amplification was performed at the genomic segments a–g as indicated in Fig. 1. Lanes W/W correspond to normal control, lanes M/M correspond to a homozygote for the major MLIV mutation and lanes D/D correspond to a homozygote of the deletion. The fragments and their sizes are indicated on the sides of the gel. Lane L is a ϕ x174 molecular weight marker.

Until now, MLIV was one of the few lysosomal storage disorders for which neither the metabolic defect nor the relevant gene had been identified; thus, no specific and accurate diagnostic tools were available. This has probably resulted in the underdiagnosis of this disease due to difficulties in accurately identifying MLIV patients in the clinic. The identification of the gene causing MLIV will now permit accurate diagnosis of this disease, including prenatal diagnosis and identification of heterozygotes. The fact that two mutations account for 95% of the MLIV alleles in the Ashkenazi population will permit effective population screening for the identification of carriers and the ascertainment of high-risk families before the birth of the first affected child.

Methods

MLIV families. We studied 24 MLIV patients diagnosed at the Hadassah University Hospital as described². These patients included 21 Ashkenazi Jews, 2 European, non-Jews and 1 Arab-Druze. In addition, their parents and 37 unaffected siblings were studied along with 60 normal controls (Ashkenazi Jews). The MLIV families were referred for diagnosis.

DNA studies. We extracted genomic DNA from peripheral blood or cultured fibroblasts by standard procedures.

Chromosome 19 genomic clones. The map and sequences of five clones in the MLIV gene region were obtained from Lawrence Livermore National Laboratory (<http://www-biol.llnl.gov/bbrp/genome/genome.html>)

Mutation detection. Genomic DNA was subjected to PCR analysis of coding regions of the three genes. The last four exons of *NTE* were found on another clone (AC009003). We designed primer pairs using Oligo primer analysis software (<http://medprobe.com/is/oligo/html>). *MCOLN1* was amplified by PCR as indicated in Fig. 1 (details and primers available on request).

All PCR reactions were performed in a thermocycler (PTC-200, MJ) with a 3-min initial denaturation and 30 cycles of denaturation (94 °C for 15 s) annealing (30 s) and elongation (72 °C) time in proportion to the size of the fragments (1 min/1 kb). We cleaned the PCR products on a 96-well Gel Filtration Block (Edge BioSystems), sequenced using dye terminators of Big-Dyes kits (Perkin-Elmer/Applied Biosystems) and analysed on an ABI 3700 sequencer (Perkin-Elmer/Applied biosystems). Sequence comparisons were performed using the STADEN package²⁰ and Sequencher software (Gene Codes).

KpnI analysis. The PCR products of fragment c were subjected to *KpnI* digestion (3 h at 37 °C) followed by agarose (2%) gel analysis.

Aberrant splicing analysis. RNA was extracted from fibroblasts using an RNA extraction kit (Qiagen). We carried out reverse transcription with AMV reverse transcriptase (Boehringer) according to the manufacturer's instructions. For RT-PCR we used primers exon 3F, 5'-TGTACCAGGC CATCTTCCAT-3', and exon10R, 5'-AGTTGTGGAAGAAGGTCAGG-3'.

MCOLN1 deletion sequencing. Using the primers C151F, 5'-CTACGCCT

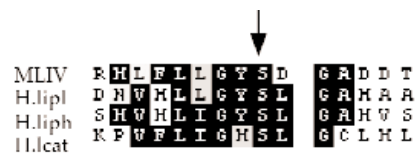


Fig. 5 Protein sequence analyses of the mucolin family. h-mucolin1, human MLIV protein; h-mucolin2, a putative protein encoded on chromosome 1p22.2–31.1 (h-mucolin3 and h-mucolin4 are two putative proteins encoded on chromosome 8 (not shown)); d-mucolin, the putative *D. melanogaster* homologue; c-mucolin, the putative *C. elegans* homologue. Multiple alignment of the mucolin1 serine-lipase PROSITE pattern region (aa 102–116) is shown with the similar motif in other human lipases: H.liplh (triacylglycerol lipase, hepatic precursor), H.lipl (lipoprotein lipase precursor) and H.lcat (phosphatidylcholine-sterol acyltransferase precursor). The consensus pattern (aa 104–113) is [LIV]-x-[LIVFY]-[LIVMST]-G-[HYWV]-s-x-G-[GSTAC]. The catalytic serine residue is indicated (arrow).

GCCCCACTG-3', and ESTeR, 5'-CCAAGGACCACAGGGGACA-3', on the genomic DNA of a patient homozygous for the deletion, we obtained a PCR product of 3,075 bp and sequenced it as above. This sequence defined the deletion borders: upstream sequence 5'-CAGCCTCGACCTCTGGGCT-3' and downstream sequence 5'-CAACAAAGCACACAGTGGGC-3'.

MCOLN1 cDNA analysis. We performed PCR of various cDNA libraries using two sets of primers: (i) exon1F, 5'-GTGACAGCGCGGGCGAT-3', and exon13R, 5'-CGCAGCAGCAGAGAAGGC-3'; (ii) exon2F, 5'-GACACCCCC AGAAGAGGAA-3', and exon14R, 5'-TCAACACCTCCCCACCA-3'. The expected lengths of 1,773 bp and 1,823 bp, respectively, were obtained in all human tissues tested (testis, cerebellum, kidney and fetal fibroblasts). The bands obtained from testis cDNA (Marathon-Ready cDNA, Clontech) were sequence verified. The primer 1F begins at nucleotide +14 of the cDNA and the primer 14R is located at the very end of the cDNA sequence; thus, the first AUG and the stop codon were included in the verified sequence.

PCR reactions were carried out with an EX-Taq PCR kit (Takara Biomedicals) with the standard buffers in the presence of 1% DMSO. PCR conditions were as follows: 94 °C for 4–7 min, followed by 40 cycles of 94 °C for 15 s, 64 °C for 30 s, 72 °C for 2 min. PCR products were cleaned and sequenced as above.

Northern-blot analysis. A human northern-blot membrane from 12 different tissues (Clontech) was hybridized with a cDNA probe of exons 1–6 of *MCOLN1* (primers EST-RTE, 5'-ATCGGACCCAGGCTGCC-3', and exon 6-RTL, 5'-AGGTTAATGGTCTCAGCCG-3'). PCR conditions were as above. We performed the hybridization procedure according to the manufacturer's instructions. The membrane was subsequently reprobated with β -actin cDNA as an internal control. Quantification of the signals was performed using the Bio-imaging Analyzer, BAS 1000 (Fujix) with Tina-2.10G program.

Web-based databases and sequence analysis. For homology searches we used BLAST (<http://www.ncbi.nlm.gov/cgi-bin/BLAST/nph-newblast?form=1>) and FASTA (<http://www2.ebi.ac.uk/fasta3/>). The putative protein was analysed by the following computer programs: EXPASY Proteomic tools contain Profile scan and PROSITE softwares (<http://www.expasy.ch/>), which were used for identification of the serine-lipase (PS00120) and the bipartite nuclear localization (PS50079) patterns; BLOCKS (<http://bioinformatics.weizmann.ac.il/blocks/>); and Kyle Doolittle Hydrophathy analysis (<http://bioinformatics.weizmann.ac.il/hydroph/>). PRSS analysis¹⁶ was performed with a window of 20 aa.

Multiple alignments were done with ClustalW (ref. 21) using the default pairwise gap opening penalty of 10 and gap extension penalty of 0.2. For EST assembly and putative gene construction, we used two tentative human gene consensus sequences (THC 330613 and THC 97830) taken from TIGR Human Gene Index (http://tigr.org/tldb/hgi/searching/hgi_reports.html) and two EST sequences (A1815981 and T66341). This was compared with the results of LabOnWeb (<http://www.labonweb.com/FloW/index.html>).

GenBank accession numbers. *MCOLN1* cDNA, AF249319; human genomic clone containing *MCOLN1*, AC008878; human mucolin2 (*MCOLN2*), AL139150; human mucolin3 (*MCOLN3*) and mucolin4 (*MCOLN4*), AC013291; *Drosophila* mucolin, AC017729; *C. elegans* mucolin, AC006679.

Acknowledgements

We thank the MLIV families for cooperation; the referring clinicians; L. Ashworth for the sequence of chromosome 19; and D. Abeliovich and I. Lerer for help. Part of this work was funded by the generous donation of P. Altura for the R. Altura Fund and by a grant from the Mucopolipidosis type IV Foundation. Part of this work was funded by an Infrastructure grant of the

Israeli Ministry of Science, the Crown Human Genome Center at The Weizmann Institute of Science and the Krupp Foundation.

Received 25 February; accepted 23 June 2000.

- Berman, E.R., Livni, N., Shapira, E. & Merin, I.S. Congenital Cornea clouding with abnormal systemic storage bodies: a new variant of Mucopolipidosis. *J. Pediatr.* **84**, 519–526 (1974).
- Amir, N., Zlotogora, J. & Bach, G. Mucopolipidosis type IV: clinical spectrum and natural history. *Pediatrics* **79**, 953–959 (1987).
- Reiss, S., Sheffer, R., Merin, S., Luder, A. & Bach, G. Mucopolipidosis type IV: a late onset and mild form. *Am. J. Med. Genet.* **47**, 392–394 (1993).
- Chitayat, D. *et al.* Mucopolipidosis type IV: clinical manifestations and natural history. *Am. J. Med. Genet.* **41**, 313–318 (1991).
- Merin, S., Livni, N., Berman, E.R. & Yatziv, S. Mucopolipidosis type IV: ocular, systemic and ultrastructural findings. *Invest. Ophthalmol.* **14**, 437–448 (1975).
- Goebel, H.H., Kohlschutter, A. & Lenard, H.G. Morphological and chemical biopsy findings in Mucopolipidosis type IV. *Clin. Neuropathol.* **1**, 73–82 (1982).
- Slaugenhaupt, S.A. *et al.* Mapping of the Mucopolipidosis type IV gene to chromosome 19p and definition of founder haplotype. *Am. J. Hum. Genet.* **65**, 773–778 (1999).
- Glusman, G. & Lancet, D. GESTALT: a workbench for automatic integration and visualization of large-scale genomic sequence analyses. *Bioinformatics* **16**, 482–483 (2000).
- Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
- Solovyev, V. & Salamov, A. The Gene-Finder, computer tools for analysis of human and model organisms genome sequences. *Ismb* **5**, 294–302 (1997).
- Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Bach, G., Zlotogora, J. & Zeigler, M. Lysosomal storage disorders among Jews. in *Genetic Diversity Among Jews* (eds Bonne-Tamir, B. & Adam, A.) 301–304 (Oxford University Press, New York, 1992).
- Kyte, J. & Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
- Rost, B. & Sander, C. Improved prediction of protein secondary structure by use of sequence profiles and neural networks. *Proc. Natl Acad. Sci USA* **90**, 7558–7562 (1993).
- Sessa, A., Ghiggeri, G.M. & Turco, A.E. Autosomal dominant polycystic kidney disease: clinical and genetic aspects. *J. Nephrol.* **10**, 295–310 (1997).
- Pearson, W.R. & Lipman, D.J. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85**, 2444–2448 (1988).
- Harris, P.C. Autosomal dominant polycystic kidney disease: clues to pathogenesis. *Hum. Mol. Genet.* **8**, 1861–1866 (1999).
- Hofmann, K., Bucher, P., Falquet, L. & Bairoch, A. The PROSITE database, its status in 1999. *Nucleic Acids Res.* **27**, 215–219 (1999).
- Gribskov, M. *et al.* Profile analysis: detection of distantly related proteins. *Proc. Natl Acad. Sci. USA* **84**, 4355–4358 (1987).
- Bonfield, J.K., Smith, K.F. & Staden, R. A new DNA sequence assembly program. *Nucleic Acids Res.* **23**, 4992–4999 (1995).
- Higgins, D.G., Thompson, J.D. & Gibson, T.J. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266**, 383–402 (1996).