

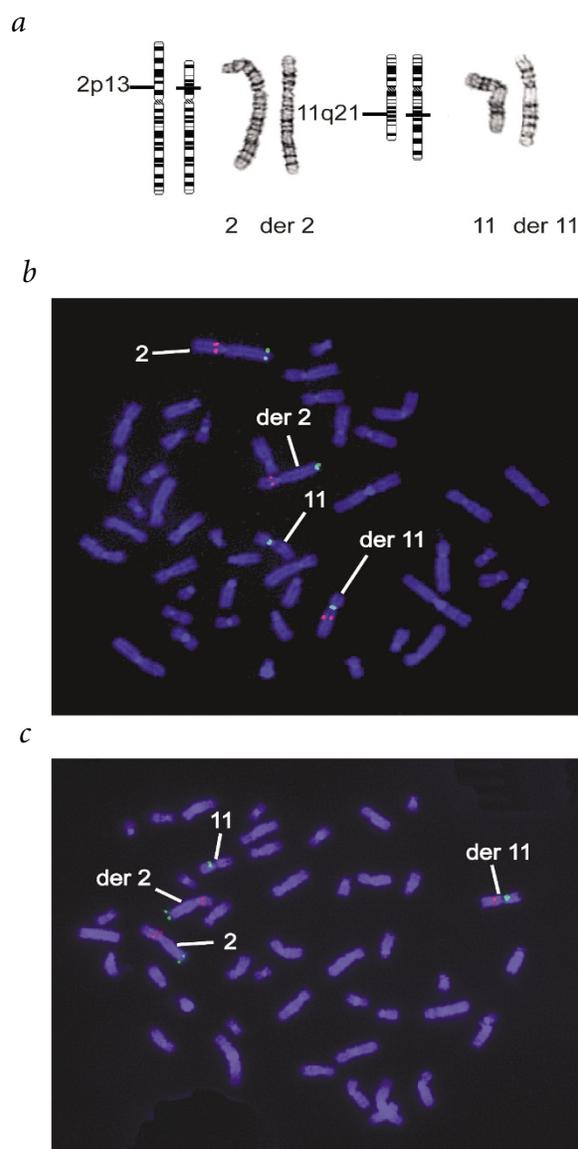
# Mutation of *ALMS1*, a large gene with a tandem repeat encoding 47 amino acids, causes Alström syndrome

Tom Hearn<sup>1</sup>, Glenn L. Renforth<sup>1</sup>, Cosma Spalluto<sup>1</sup>, Neil A. Hanley<sup>1</sup>, Karen Piper<sup>1</sup>, Sarah Brickwood<sup>1</sup>, Chris White<sup>2</sup>, Vincent Connolly<sup>3</sup>, James F.N. Taylor<sup>4</sup>, Isabelle Russell-Eggitt<sup>4</sup>, Dominique Bonneau<sup>5</sup>, Mark Walker<sup>2</sup> & David I. Wilson<sup>1</sup>

Published online: 8 April 2002, DOI: 10.1038/ng874

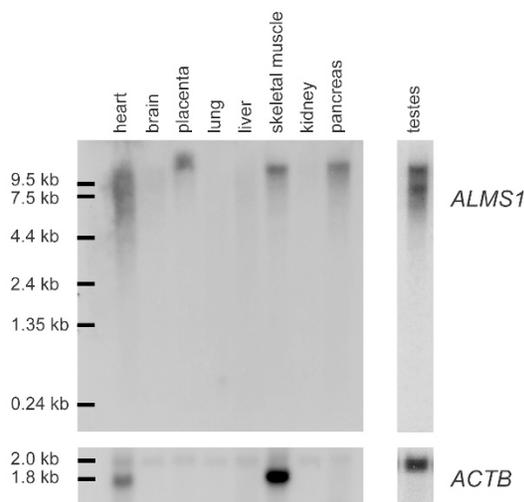
Alström syndrome (OMIM 203800) is an autosomal recessive disease, characterized by cone-rod retinal dystrophy, cardiomyopathy and type 2 diabetes mellitus, that has been mapped to chromosome 2p13 (refs 1–5). We have studied an individual with Alström syndrome carrying a familial balanced reciprocal chromosome translocation (46, XY,t(2;11)(p13;q21)mat) involving the previously implicated critical region. We postulated that this individual was a compound heterozygote, carrying one copy of a gene disrupted by the translocation and the other copy disrupted by an intragenic mutation. We mapped the 2p13 breakpoint on the maternal allele to a genomic fragment of 1.7 kb which contains exon 4 and the start of exon 5 of a newly discovered gene (*ALMS1*); we detected a frameshift mutation in the paternal copy of the gene. The 12.9-kb transcript of *ALMS1* encodes a protein of 4,169 amino acids whose function is unknown. The protein contains a large tandem-repeat domain comprising 34 imperfect repetitions of 47 amino acids. We have detected six different mutations (two nonsense and four frameshift mutations causing premature stop codons) in seven families, confirming that *ALMS1* is the gene underlying Alström syndrome. We believe that *ALMS1* is the first human disease gene characterized by autosomal recessive inheritance to be identified as a result of a balanced reciprocal translocation.

Alström syndrome was initially mapped to an interval of 6.1 cM between loci *D2S286* and *D2S327* (refs 3–5). Although several candidate genes have been investigated, no mutations have previously been identified<sup>6–8</sup>. We have shown that the 2p13 breakpoint in the individual with the 46,XY,t(2;11)(p13;q21)mat translocation is between these loci by metaphase fluorescence *in situ* hybridization (FISH) analysis using the BACs RP11-355F16 (containing *D2S286*) and RP11-480F1 (located 150 kb proximal to *D2S327*) as probes. The BAC RP11-582H21 crosses the translocation breakpoint (Fig. 1a,b) and is overlapped by RP11-79N18, which contains *CCT7*, a member of a chaperonin gene



**Fig. 1** Partial karyotype and FISH analysis of 46, XY,t(2;11)(p13;q21). **a**, Partial ideograms and karyotypes showing translocation breakpoints. **b**, BAC RP11-582H21 (red) hybridizes to chromosomes 2, derivative 2 (der2) and derivative 11 (der11). **c**, Long-range PCR product of 9 kb (red) from within RP11-582H21 that contains *ALMS1* exons 3–5 and that also hybridizes to chromosomes 2, der 2 and der 11, and thus crosses the 2p13 breakpoint. Chromosome 2 PAC RP5-1011O17 (green) and chromosome 11-specific alphoid DNA (green) were used as control probes.

<sup>1</sup>Division of Human Genetics, Southampton University, The Duthie Building, Southampton General Hospital, Tremona Road, Southampton SO16 6YD, UK. <sup>2</sup>Department of Diabetes and Metabolism, The Medical School, Framlington Place, University of Newcastle, UK. <sup>3</sup>South Cleveland Hospital, Middlesbrough, UK. <sup>4</sup>Hospital for Sick Children, Great Ormond Street, London UK. <sup>5</sup>Department of Medical Genetics, Centre Hospitalier Universitaire, Angers, France. Correspondence should be addressed to D.I.W. (d.i.wilson@soton.ac.uk) or M.W. (mark.walker@ncl.ac.uk).



**Fig. 2** Multiple-tissue northern blot showing *ALMS1* expression. Hybridization was carried out with probes consisting of the PCR product of *ALMS1* (3,651–4,671 bp of the coding sequence) and then with *ACTB* as a control.

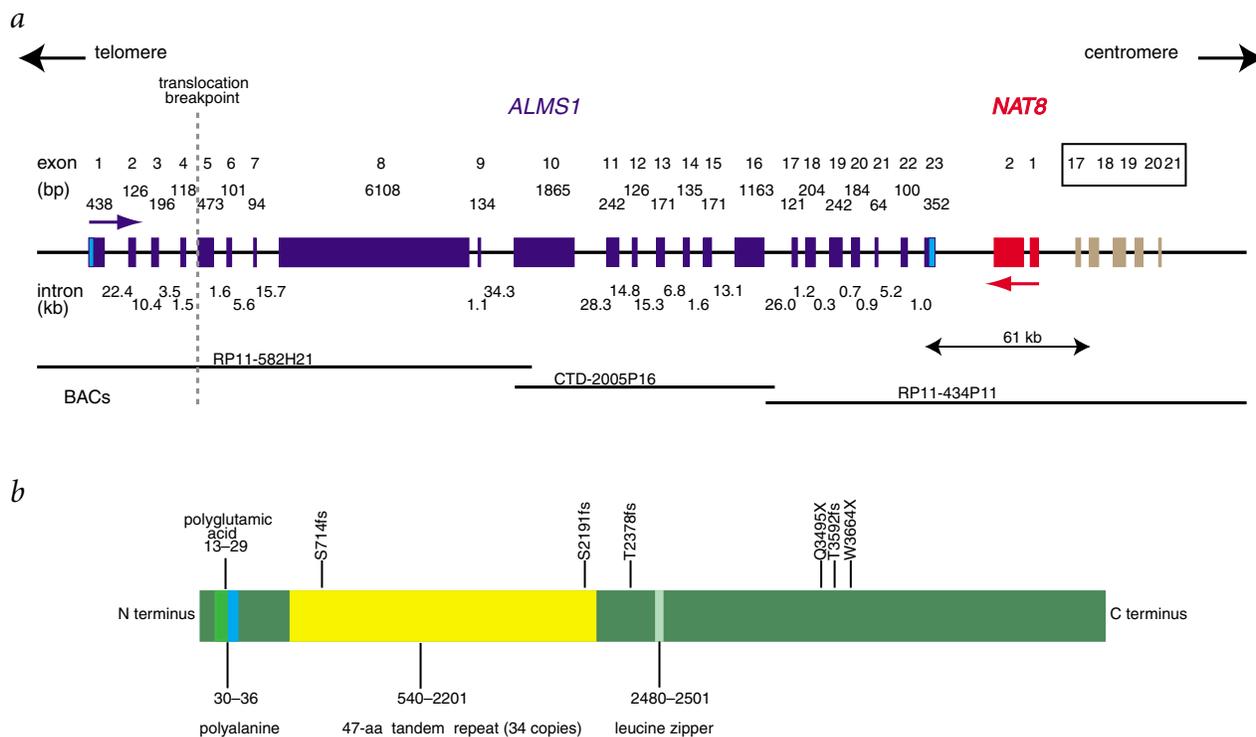
family<sup>9</sup>. Mutations in *MKKS* (encoding a putative chaperonin) have been detected in Bardet-Biedl syndrome<sup>10,11</sup>, which has phenotypic overlap with Alström syndrome (including obesity, insulin resistance and retinopathy). We therefore examined individuals with Alström syndrome for mutations in *CCT7*, but did not identify any changes in coding sequence.

Sequence analysis of RP11-582H21 suggested the presence of a single gene represented by the cDNA KIAA0328 (ref. 12). Northern-blot analysis using an RT-PCR product of 3.9 kb derived from this cDNA disclosed a transcript of 12–13 kb, suggesting that the 6.3-kb KIAA0328 database entry was a partial sequence.

We used RT-PCR with exon prediction and specific primer design to identify a longer transcript that was confirmed by northern-blot analysis (Fig. 2). The extended sequence (*ALMS1*) is 12.9 kb in length and contains an open reading frame of 12.5 kb comprising 23 exons (Fig. 3a) and over 224 kb of genomic DNA (although the most abundant transcript detected by PCR does not contain exon 2). Analysis of the *ALMS1* transcript revealed the predicted start codon in an efficient Kozak site with suitable transcription initiation less than 100 bp upstream of the 5' limit of our cDNA sequence.

A duplication of the 6.5-kb region corresponding to exons 17–21 lies 61 kb downstream of *ALMS1*. The region shares 94.5% sequence identity and is in the same orientation; however, the presence of frameshift mutations indicates that the duplicated versions of exons 17–21 are not expressed. We designed PCR primers so as to screen the coding copies of exons 17–21 only. In between *ALMS1* and the duplicated region is *NAT8*, a gene on the opposite strand that encodes a product similar to bacterial acetyltransferases<sup>13</sup>.

We localized the maternally inherited 2p13 breakpoint to a long-range PCR product of 9.0 kb (Fig. 1c), derived from RP11-582H21, containing exons 3–5. We refined the localization to an *EcoRI* fragment of 1.7 kb containing exon 4 and the start of exon 5 by Southern-blot analysis. In the same individual, we also detected a deletion of 2 bp in exon 8 of the paternal *ALMS1* allele (2141delCT) that is predicted to cause premature termination five codons downstream of the deletion. In agreement



**Fig. 3** Structure of *ALMS1* and protein. **a**, Intron and exon organization (not to scale). *ALMS1* exons are shown in dark blue, untranslated regions are shown in light blue and the copies of exons 17–21 in the duplicated region are shown in brown. *NAT8* exons are shown in red and are in the opposite orientation to *ALMS1* (red and blue arrows). The position of the 2p13 translocation breakpoint is indicated by a dotted line. The genomic BAC contig is shown beneath. **b**, Primary structure of *ALMS1* and position of premature stop codons causing protein truncation.

Table 1 • Families investigated and mutations detected

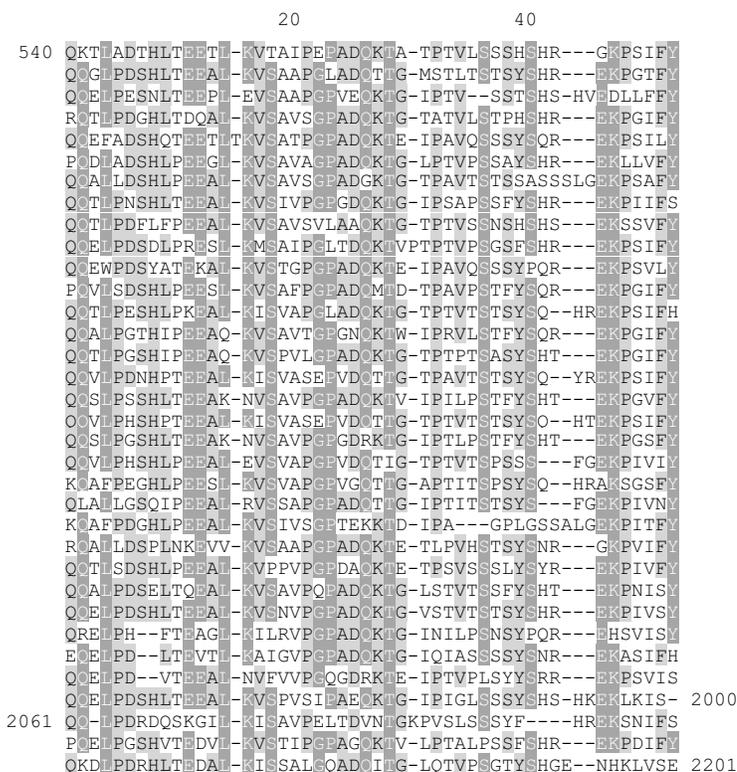
|                | Status           | DNA mutation                | Protein change        | Comments                       |
|----------------|------------------|-----------------------------|-----------------------|--------------------------------|
| Family 1       |                  |                             |                       |                                |
| F1M            | carrier (mother) | translocation               |                       | 46, XX,t(2;11)(p13;q21)        |
| F1F            | carrier (father) | 2141delCT                   | Ser714fs              |                                |
| F1Ch           | affected         | translocation / 2141delCT   | Ser714fs              | 46, XY,t(2;11)(p13;q21)        |
| Family 2       |                  |                             |                       |                                |
| F2F (BV0757)   | carrier (father) | 10775delC                   | Thr3592fs             | no DNA from mother             |
| F2Ch1 (BV0791) | affected         | [10775delC] + [10992G→A]    | Thr3592fs + Trp3664X  |                                |
| F2Ch2 (BV0792) | affected         | [10775delC] + [10992G→A]    | Thr3592fs + Trp3664X  |                                |
| Family 3       |                  |                             |                       |                                |
| F3M            | carrier (mother) | 10775delC                   | Thr3592fs             |                                |
| F3F            | carrier (father) | no coding changes           |                       |                                |
| F3Ch           | affected         | [10775delC] + [?]           | Thr3592fs + ?         | paternal mutation unidentified |
| Family 4       |                  |                             |                       |                                |
| F4Ch           | affected         | [10775delC] + [10775delC]   | Pro3592fs + Thr3592fs | no parental DNA                |
| Family 5       |                  |                             |                       |                                |
| F5F (BV0764)   | carrier (father) | 10775delC                   | Thr3592fs             | no DNA from mother             |
| F5Ch           | affected         | [10775delC] + [6571delTCAC] | Thr3592fs + Ser2191fs |                                |
| Family 6       |                  |                             |                       |                                |
| F6Ch           | affected         | [10483C→T] + [10483C→T]     | Gln3495X + Gln3495X   | no parental DNA                |
| Family 7       |                  |                             |                       |                                |
| F7Ch           | affected         | [7132insA] + [7132insA]     | Thr2378fs             | no parental DNA                |
| Family 8       |                  |                             |                       |                                |
| F8M (BV0752)   | carrier (mother) | no coding changes           |                       | no DNA from affected child     |
| F8F (BV0754)   | carrier (father) | no coding changes           |                       |                                |
| Family 9       |                  |                             |                       |                                |
| F9Ch           | affected         | no coding changes           |                       | no parental DNA                |

M, mother; F, father; Ch, child.

with these findings, sequencing of an RT-PCR product derived from lymphocyte mRNA showed expression of only a transcript containing the 2-bp deletion (2141delCU) mutation, consistent with monoallelic expression of the paternal allele and disruption of the maternal *ALMS1* transcript between exons 4 and 5.

We studied six other families with Alström syndrome (having either affected individuals or obligate gene carriers) and identified five additional mutations, each causing a premature stop codon (see Table 1 and Web Fig. A online). In families 8 and 9, we were unable to identify pathogenic mutations in coding sequences. We identified two mutations in exon 16 of family 2. The affected siblings are compound heterozygotes harboring a nonsense mutation (Trp3664X) and a single-base pair deletion (10775delC) resulting in a predicted frameshift causing premature termination five codons downstream (Thr3592fs). DNA from the father (F2F) was heterozygous with respect to the 10775delC mutation, but lacked the Trp3664X mutation, consistent with carrier status. Maternal DNA was unavailable. Sequencing of RT-PCR products from lymphocyte mRNA revealed expression of the respective

mutations in the father and children. We also detected the 10775delC mutation in three additional families not known to be related to family 2. In family 3, the affected individual



**Fig. 4** Tandem amino-acid repeat. The sequence alignment of 34 copies of a tandem-repeat unit identified in *ALMS1* (residues 540–2201). The repeat unit is not perfect (length 45–50 aa and identity 40–90%). The array is interrupted by a region of 60 aa composed of incomplete degenerate repeats (residues 2001–2060), which for purposes of alignment is not shown. The array is encoded entirely by exon 8 and contains no cysteine residues. Light and dark shading indicate greater than 60% and greater than 80% conservation, respectively.



inherited the 10775delC mutation from his mother; sequencing the entire coding and splice-site regions failed to detect a paternally inherited mutation. We identified a paternally inherited, heterozygous A→G transition in exon 8, causing a substitution of aspartic acid for asparagine (Asn1788Asp); however, we also found this change in 3 of 72 normal chromosomes, suggesting that it is not pathogenic.

We identified mutations in three additional affected individuals. One was homozygous with respect to a nonsense mutation in exon 16 (Gln3495X). The second was homozygous with respect to an insertion of a single base pair in exon 8 (7132insA), causing a frameshift mutation (Thr2378fs). The third had a deletion of 4 bp in exon 8 (6571delTCAC), causing a frameshift mutation (Ser219fs), in addition to the 10775delC mutation. None of the truncating mutations identified in this study were found in 100 normal chromosomes. These data indicate that *ALMS1* is the gene underlying Alström syndrome.

The *ALMS1* protein consists of 4,169 aa and has a predicted molecular weight of 461.2 kD. *In silico* analysis predicts a leucine zipper at aa 2480–2501 and a potential signal peptide at aa 211–223, but no other known evolutionarily conserved sequence domains are apparent (Fig. 3b). A striking feature of the sequence is a large tandem-repeat domain comprising 34 imperfect repetitions of 47 aa (aa 540–2201; Fig. 4) that contains no cysteine residues. This array constitutes approximately 40% of the protein and is encoded entirely by exon 8. Repeat length varies from 45 to 50 aa, and repeat identity ranges from 40% to 90%. *ALMS1* also contains a run of 17 glutamic-acid residues (aa 13–29) encoded by (GAG)<sub>13</sub>GAA(GAG)<sub>3</sub>, followed by a run of 7 alanine residues (aa 30–36).

To identify other potential pathogenetic mechanisms, we analyzed the GAG trinucleotide repeat and polyalanine tract in exon 1 for size variation. Sequencing of 72 normal chromosomes revealed a size variation of 12–20 glutamic acids; however, the GAG allele sizes of the obligate carriers (F3F, F8M, F8F) and affected individuals (F9Ch), in whom no other mutations have been detected, lay within the normal variation range. We did not find any variation in the size of the polyalanine tract in either normal individuals or obligate gene carriers.

Using RT-PCR, we demonstrated *ALMS1* expression in all of the tissues and cells investigated: fetal heart, aorta, liver, kidney, lung, neural tube, eye, adrenal glands, placenta, lymphocytes and WERI retinoblastoma cell line.

Cloning breakpoints of balanced translocations has proved a successful strategy for identifying dominant and X-linked, but not autosomal recessive, disease genes. We believe that *ALMS1* is the first human autosomal gene involved in recessive disease to be identified using this strategy. The six independent mutations predicted to cause protein truncation and the accompanying report by Collin *et al.*<sup>14</sup> confirm that dysfunction of *ALMS1* causes Alström syndrome. We failed to identify mutations in two families, which raises the possibility of genetic heterogeneity, although our methods of mutation detection did not exclude larger DNA deletions or mutations within regulatory regions.

The function of *ALMS1* is not clear. There are similarities between the structural organization of *ALMS1* and that of mucin genes<sup>15</sup>. Mucins are secreted proteins that are heavily glycosylated and have a large tandem-repeat domain encoded by a single exon. Like *ALMS1* repeats, mucin tandem repeats have a low cysteine-residue content, but they also have characteristic high threonine and serine content<sup>15</sup> not found in *ALMS1* repeats, suggesting that *ALMS1* is not a mucin. *ALMS1* is a new protein involved in an insulin-resistance syndrome and thus represents the potential for discovery of a new pathway.

## Methods

**Patient samples.** We obtained samples from nine families with Alström syndrome, including lymphoblastoid cell lines from members of three families from the European Collection of Cell Cultures (Salisbury, UK; BV0752, BV0754, BV0757, BV0791, BV0792, BV0764). In total, we used DNA from nine affected children; in two families, DNA was available from only one parent. We used the DNA of 50 unrelated normal individuals as controls. Ethical approval was obtained from the Newcastle Local Research Ethics committee.

**Sequence analysis.** We identified BACs within the 6.1-cM critical region using BLAST to query the nr and htgs databases with marker sequences. We then built BAC contigs electronically using the TIGR (The Institute for Genomic Research) BAC end-sequence database and NIX analysis. We identified genes using NIX and analyzed protein sequences using PIX and InterProScan.

**Mutation detection.** We sequenced PCR products using a BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) according to the manufacturer's instructions and analyzed the reaction products on an ABI 377 sequencer. We designed primers to amplify exon and splice-site sequences from genomic DNA.

**FISH analysis.** We labeled probes for FISH following the manufacturer's instructions (Vysis) and hybridized them to metaphase chromosomes. We obtained RP11 BACs from BAC/PAC Resources and CTD-2005P16 from Research Genetics. We generated long-range PCR products using an Expand Long Template PCR kit (Roche).

**RT-PCR and RACE.** We extracted total RNA from lymphoblastoid cells using Tri reagent (Sigma) and generated first-strand cDNA using Superscript II (Invitrogen). We carried out 5' RACE as described<sup>16</sup>. We obtained fetal tissues with approval from the Southampton and South West Hampshire Joint Local Research Ethics Committee and after receiving informed consent.

**Southern-blot analysis.** We separated restriction digests of total genomic DNA on a 1% agarose gel and transferred them to membrane by alkali blotting. We purified a 346-bp PCR product spanning part of exon 4 and part of the following intron by gel extraction and labeled it with [ $\alpha$ -<sup>32</sup>P]dCTP using a Rediprime II kit (Amersham Pharmacia Biotech). We followed a standard filter hybridization and washing protocol.

**Northern-blot analysis.** We hybridized human multiple-tissue northern blots (Clontech) with a probe of either a 3.9-kb RT-PCR product spanning exons 10–20 (8,244–12,161 bp of the *ALMS1* coding sequence) or the PCR product of Ex8fF/Ex8fR (3,651–4,671 bp) and then with ACTB as a control. We purified and labeled the probes as described above for Southern-blot analysis and followed the manufacturer's instructions for hybridization and washing conditions (Clontech).

**GenBank accession number.** *ALMS1* mRNA sequence, AJ417593.

*Note: Supplementary information is available on the Nature Genetics website.*

## Acknowledgments

We are grateful to M. Rocchi for the PAC 1011O17 and chromosome 11 alphoid FISH probes. We wish to thank D. Wellesley, D. Schapira, A. Moore, S. O'Rahilly, I. Ellis and N. Roper for help ascertaining families. We are very grateful for the help of families and individuals with Alström syndrome and the support of the Juvenile Diabetes Research Foundation (to K.P. and S.B.). N.A.H. is a Department of Health (UK) Clinician Scientist. This study was funded by Diabetes UK.

## Competing interests statement

The authors declare competing financial interests: see the Nature Genetics website (<http://genetics.nature.com>) for details.

Received 24 October 2001; accepted 28 January 2002.

1. Russell-Eggitt, I.M. *et al.* Alström syndrome. Report of 22 cases and literature review. *Ophthalmology* **105**, 1274–1280 (1998).
2. Marshall, J.D. *et al.* Genealogy, natural history, and phenotype of Alström syndrome in a large Acadian kindred and three additional families. *Am. J. Med. Genet.* **73**, 150–161 (1997).
3. Collin, G.B., Marshall, J.D., Cardon, L.R. & Nishina, P.M. Homozygosity mapping at Alström syndrome to chromosome 2p. *Hum. Mol. Genet.* **6**, 213–219 (1997).
4. Macari, F. *et al.* Refinement of genetic localization of the Alström syndrome on chromosome 2p12-13 by linkage analysis in a North African family. *Hum. Genet.* **103**, 658–661 (1998).
5. Collin, G.B. *et al.* Alström syndrome: further evidence for linkage to human chromosome 2p13. *Hum. Genet.* **105**, 474–479 (1999).
6. Collin, G.B., Nishina, P.M., Marshall, J.D. & Naggert, J.K. Human DCTN1: genomic structure and evaluation as a candidate for Alström syndrome. *Genomics* **53**, 359–364 (1998).
7. Collin, G.B., Marshall, J.D., Naggert, J.K. & Nishina P.M. TGFA: exon-intron structure and evaluation as a candidate gene for Alström syndrome. *Clin. Genet.* **55**, 61–62 (1999).
8. Chen, F. *et al.* Characterization of the murine Lbx2 promoter, identification of the human homologue, and evaluation as a candidate for Alström syndrome. *Genomics* **74**, 219–227 (2001).
9. Leroux, M.R. & Hartl, F.U. Protein folding: versatility of the cytosolic chaperonin TRiC/CCT. *Curr. Biol.* **10**, R260–264 (2000).
10. Katsanis, N. *et al.* Mutations in MKKS cause obesity, retinal dystrophy and renal malformations associated with Bardet-Biedl syndrome. *Nature Genet.* **26**, 67–70 (2000).
11. Slavotinek, A. *et al.* Mutations in MKKS cause Bardet-Biedl syndrome. *Nature Genet.* **26**, 15–16 (2000).
12. Nagase, T. *et al.* Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins *in vitro*. *DNA Res.* **4**, 141–150 (1997).
13. Ozaki, K., Fujiwara, T., Nakamura, Y. & Takahashi, E. Isolation and mapping of a novel human kidney- and liver-specific gene homologous to the bacterial acetyltransferases. *J. Hum. Genet.* **43**, 255–258 (1998).
14. Collin, G.B. *et al.* Identification of a novel gene causing obesity, type 2 diabetes and neurosensory degeneration in Alström syndrome. *Nature Genet.* **31** (2002); advance online publication, (DOI: 10.1038/ng867).
15. Moniaux, N., Escande, F., Porchet, N., Aubert, P. & Batra, S.K. Structural organization and classification of the human mucin genes. *Front. Biosci.* **6**, 1192–1206 (2001).
16. Matz, M. *et al.* Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res.* **27**, 1558–1560 (1999).