Mutations in a new gene, encoding a zinc-finger protein, cause tricho-rhino-phalangeal syndrome type I

Parastoo Momeni[1]*, Gernot Glöckner[2]*, Olaf Schmidt[1], Diane von Holtum[1], Beate Albrecht[1], Gabriele Gillessen-Kaesbach[1], Raoul Hennekam[3], Peter Meinecke[4], Bernhard Zabel[5], André Rosenthal[2], Bernhard Horsthemke[1] & Hermann-Josef Lüdecke[1]

*The first two authors contributed equally to this work.

Tricho-rhino-phalangeal syndrome type I (TRPS I, MIM 190350) is a malformation syndrome characterized by craniofacial and skeletal abnormalities and is inherited in an autosomal dominant manner[1]. TRPS I patients have sparse scalp hair, a bulbous tip of the nose, a long flat philtrum, a thin upper vermilion border and protruding ears. Skeletal abnormalities include cone-shaped epiphyses at the phalanges, hip malformations and short stature. We assigned *TRPS1* to human chromosome 8q24. It maps proximal of *EXT1*, which is affected in a subgroup of patients with multiple cartilaginous exostoses and deleted in all patients with TRPS type II (TRPS II, or Langer-Giedion syndrome, MIM 150230; refs 2–5). We have positionally cloned a gene that spans the chromosomal breakpoint of two patients with TRPS I and is deleted in five patients with TRPS I and an interstitial deletion[4,6]. Northern-blot analyses revealed transcripts of 7 and 10.5 kb. *TRPS1* has seven exons and an ORF of 3,843 bp. The predicted protein sequence has two potential nuclear localization signals and an unusual combination of different zinc-finger motifs, including IKAROS-like and GATA-binding sequences. We identified six different nonsense mutations in ten unrelated patients. Our findings suggest that haploinsufficiency for this putative transcription factor causes TRPS I.

We have recently refined the proximal border of the minimal *TRPS1* region by mapping the inversion breakpoint near to *D8S98* in patient HB11480 (Fig. 1*a*,*b*; ref. 4). DNA sequence analysis of 3 overlapping PAC clones spanning this breakpoint and computer-based searches for ORFs revealed ORFs of 604, 929 and 1,130 bp in PAC24 and ORFs of 123 and 1,059 bp in

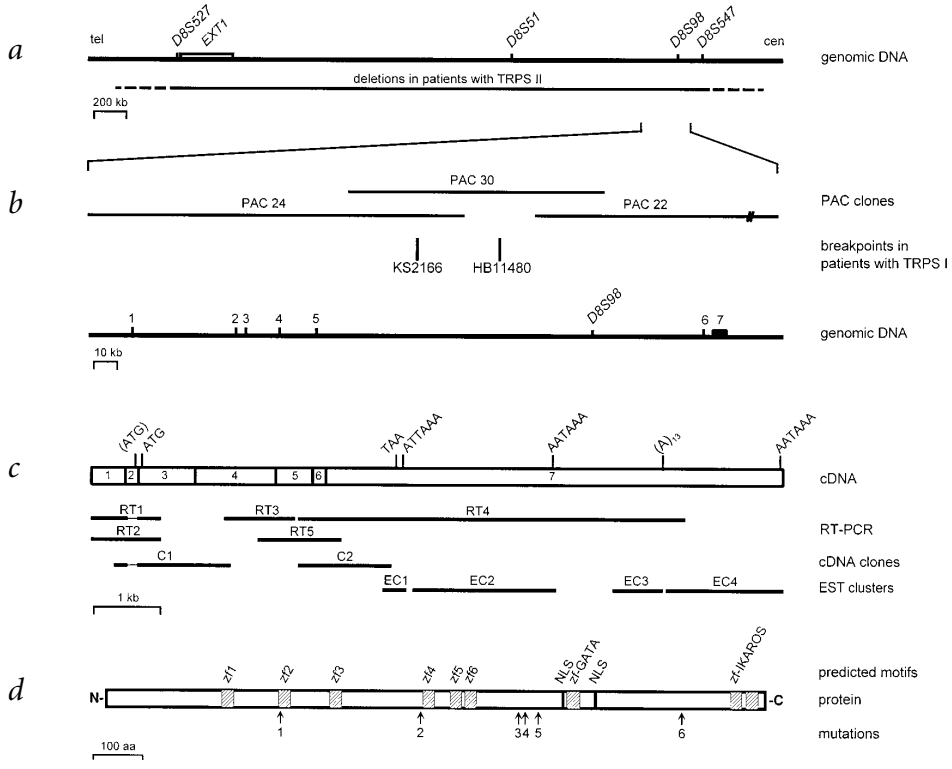

Fig. 1 Physical map of the TRPS region on human chromosome 8q24 and structure of *TRPS1* and TRPS1 protein. ***a***, Overview of the region. *EXT1*, selected DNA markers and the region commonly deleted in patients with TRPS II are indicated. tel, telomere; cen, centromere. ***b***, Genomic structure of *TRPS1*. The PAC clones 24, 30 and 22 span the gene. The approximate locations of the chromosomal breakpoints of the patients KS2166 and HB11480 are indicated by vertical lines. The seven exons of the gene are indicated (filled boxes). *D8S98* maps within intron 5. ***c***, Structure of *TRPS1* cDNA. The open boxes represent the exons. The start and stop codons, three polyadenylation signals and a poly(dA)-stretch in the 3′ UTR are indicated. A probably silent translation start codon is indicated in brackets. The cDNA is represented by the RT–PCR products and cDNA clones RT2, C1, RT3, RT5, RT4 and EC4, which are derived from fetal and infant (EC4) brain RNA. The cDNA clones C1 and C2 were isolated with probes 1 and 2. The narrow lines in RT1 and C1 indicate the absence of exon 2 in the sequences. Each EC line represents a cluster of overlapping EST clones. ***d***, Structure of the TRPS1 protein and location of mutations. Hatched boxes, zinc-finger motifs (zf). Vertical bars, putative nuclear localization signals (NLS). The location of the GATA and IKAROS-like zinc fingers are given. The numbers 1–6 indicate the location of the mutations given in Table 1.

[1]*Institut für Humangenetik, Universitätsklinikum, Essen, Germany.* [2]*Abteilung Genomanalyse, Institut für Molekulare Biotechnologie, Jena, Germany.* [3]*Department of Clinical Genetics and Pediatrics, University of Amsterdam, Amsterdam, The Netherlands.* [4]*Abteilung Medizinische Genetik, Altonaer Kinderkrankenhaus, Hamburg, Germany.* [5]*Kinderklinik und Poliklinik, Klinikum der Johannes Gutenberg-Universität Mainz, Mainz, Germany. Correspondence should be addressed to H.-J.L. (e-mail: hj.luedecke@uni-essen.de).*

Table 1 • *TRPS1* mutations in patients with TRPS I

No.[a]	Case ID	Familial	Location	Mutation[b]	Amino acid change[c]
1	6820	yes	exon 4	1014C→A	C338X
2	G1116	yes	exon 4	1831C→T	R611X
3	9438	yes	exon 5	2406–2407insG	frameshift from codon 803
4	5650	no	exon 5	2441–2442insT	frameshift from codon 814
5	1277	no	exon 5	2518C→T	R840X
6	11049	no	exon 7	3360–3361insGGAG	frameshift from codon 1121

[a]Numbers refer to the mutations shown in Fig. 1*d*. [b]Nucleotide numbers refer to *TRPS1* cDNA sequence. [c]Amino acid numbers refer to deduced peptide sequence.

PAC22 (Fig. 1*b*). The first four ORFs were flanked by two splice sites, and the 1,059-bp ORF was flanked by a potential splice-acceptor site and three in-frame translation stop codons, indicating a potential 3′ exon. Database searches with this ORF and sequences located more 3′ identified four clusters of evolutionarily conserved expressed sequence tag (EST) clones (EC1–4, Fig. 1*c*), which are scattered over 6 kb of genomic DNA. EC1 overlapped with the 1,059-bp ORF, and sequence analyses of the other three EST clusters showed complete collinearity with genomic DNA without any significant ORF. The poly(A)-tail of EC1 starts 29 bp after a non-classical (ATTAAA) polyadenylation signal. Clones of cluster EC3 have been primed at an internal $(A)_{13}$ stretch (Fig. 1*c*). Clusters EC2 and EC4 contain classical polyadenylation signals (AATAAA) preceding their poly(dA)-tails. By cDNA cloning and exon connecting PCR following reverse transcription of human fetal brain RNA (RT–PCR; Fig. 1*c*), we found that all ORFs and EST clusters belong to the same transcript, established their order and identified two non-coding 5′ exons, one of which (exon 2, 158 bp) is subject to alternative splicing. The complete cDNA sequence is 10,011 bp.

We determined the genomic structure of the putative *TRPS1* by comparing the cDNA and genomic DNA sequences. *TRPS1* is composed of seven exons, spans 260,500 bp of genomic DNA and is transcribed from telomere to centromere. The exact transcription start site, and hence the exact size of the first exon, have not yet been determined. The ORF (3,843 bp) starts at the third nucleotide of exon 3 (http://www.uni-essen.de/~thg020/). The translation start site conforms to the Kozak consensus sequence[7], regardless of whether exon 2 is used. The alternative second exon has an additional in-frame translation initiation signal, the use of which would add 13 amino acids to the amino terminus of the protein. This signal conforms poorly to the Kozak consensus sequence, however, and is therefore unlikely to be used[7]. The large, 5,530-bp 3′ UTR contains 16 copies of the AUUUA motif. This motif is known to be the recognition signal for mRNA degradation by proteosomal endonuclease activity[8], suggesting that the amount of TRPS1 protein may be regulated by cellular concentration of *TRPS1* mRNA.

Northern-blot analyses with probe 2 identified two transcripts of approximately 7 and 10.5 kb in RNA from human fetal brain, lung and kidney, but not in fetal liver (Fig. 2) or adult heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas (data not shown). The presence of EST clones from adult prostate, stomach, breast and infant brain suggests that a low level of expression occurs in these tissues. The transcript sizes correspond to the positions of the alternative polyadenylation signals used in EC2 and EC4 (Fig. 1*c*).

The ORF predicts a weakly basic (pI=7.495) polypeptide of 1,281 amino acids (http://www.uni-essen.de/~thg020/) with a calculated molecular mass of 141,580 daltons. The analysis of the amino acid sequence and the results of BlastP database searches using the BLOSUM62 and PAM30 matrices predict that *TRPS1* encodes a novel transcription factor with an unusual composition of nine putative zinc-finger motifs of four different types[9]. The first three motifs, of the C_2H_2 type (zf1–3, Fig. 1*d*), share a consensus sequence of '$C_2X_{14}H_2$' without any similarity in protein databases. Motifs 4–6 (zf4–6, Fig. 1*d*) share the consensus '$C_2X_{12}H_2$'. Database searches revealed a few similarities between motifs 4, 6 and zinc-finger DNA-binding domains of other proteins. The seventh motif is of the C_2C_2 type ('$CXNCX_{17}CNXC$'). Database searches revealed 164 'hits', exclusively with members of the family of GATA DNA-binding transcription factors. The two C_2H_2-type zinc fingers at the carboxy terminus are similar to 285 other zinc-finger motifs of proteins from many different organisms. The highest degree of similarity was found with the two characteristic C-terminal zinc fingers of the IKAROS transcription factor family (human, 52% identity and 72% similarity in 53 aa). This domain also has similarity to zinc-finger domains of the

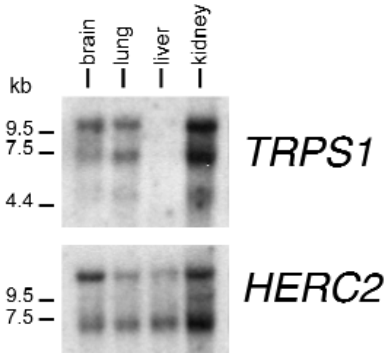

Fig. 2 Northern-blot analysis. We probed a northern blot containing ~2 µg purified poly(A)+-RNA from four human fetal tissues with radioactively labelled probe 2 (PM67/PM66). Length standards are as indicated by the supplier. Integrity of the RNA was determined by hybridization with a *HERC2* probe[19] (provided by K. Buiting).

Fig. 3 Index patient (front) of case 9438 with his mother (left) and his maternal grandmother (right).

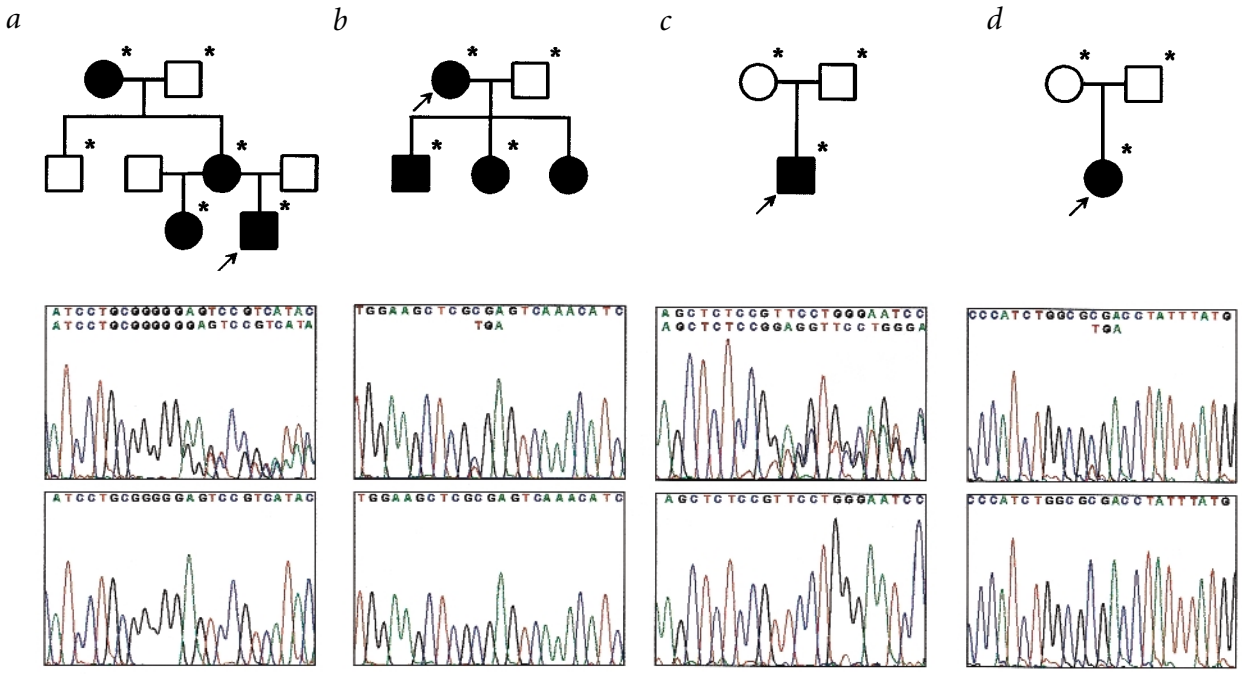

Fig. 4 Mutation analyses. The top panels under each pedigree show the electropherograms of the sequencing products obtained in the index patient (arrow). The upper line is the normal sequence and the lower line is the mutant sequence. The bottom panels show the relevant part of the sequence in a normal control individual. ***a***, One-bp insertion in case 9438. ***b***, C→T transition in case G1116. ***c***, Four-bp insertion in case 11049. ***d***, C→T transition in case 1277. Individuals available for analyses are indicated (asterisk).

Drosophila melanogaster proteins Hunchback, Krüppel and Spalt-Major, which are responsible for setting up patterns of gene expression[10]. The double zinc-finger domain of IKAROS is known to mediate protein-protein interactions[11]. In mice, Ikaros protein complexes have been found at centromeric heterochromatin associated with transcriptionally silent genes, indicating a role for these complexes in transcriptional repression[10].

The putative GATA-binding zinc finger is flanked by two potential nuclear localization signals (NLS) at positions 886–891 (LRRRRG) and 946–952 (RRRTRKR). Such strongly basic hexa/heptapeptides are highly characteristic for large proteins that function as transcription factors[12].

To search for *TRPS1* mutations, we selected ten unrelated patients with TRPS I and apparently normal chromosomes as determined by karyotype analysis (450-band stage) and fluorescence *in situ* hybridizations with YAC A3B7 (ref. 2). All patients had the typical phenotype of TRPS I as shown for case 9438 (Fig. 3). So far we have found a mutation in three familial and three sporadic cases. Three mutations are nonsense mutations, and three are insertional frameshift mutations (Table 1 and Figs 1*d* and 4*a–d*).

We propose that the described gene is *TRPS1*. First, it is deleted in all patients with TRPS I and a chromosomal deletion as well as in all patients with TRPS II. This also identifies TRPS II as a true contiguous gene syndrome. Second, *TRPS1* is disrupted by balanced chromosomal abnormalities in two patients with TRPS I (KS2166 and HB11480, Fig. 1*b*). Third, it is inactivated by mutations in three familial and three sporadic cases. All mutations result in the loss of one functional copy of *TRPS1*, presumably reducing the concentration of TRPS1 protein. The finding of haploinsufficiency of a zinc-finger transcription factor gene is not uncommon in genetic diseases. For example, hemizygous *GATA4* deletions have been found in patients with congenital heart disease[13], and inactivating *GLI3* mutations lead to Greig cephalopolysyndactyly syndrome[14]. It is possible that transcription factors such as TRPS1 are present in limited amounts and are dosage-sensitive due to altered stoichiometry in heterodimers[15]. Functional analysis of *TRPS1* will enhance our understanding of craniofacial and limb development and may provide a clue to understanding some of the phenotypic diversity of the human face.

Methods

Sources of genomic and EST clones. We isolated bacteriophage P1-derived artificial chromosome clones (PACs) from the human PAC library RPCI1, 3-5 (ref. 16). We obtained filters and clones from the Resource Centre of the German Human Genome Project (RZPD) at the Max-Planck-Institute for Molecular Genetics, and EST clones from either the RZPD or the UK HGMP Resource Centre. We determined the insert sequences of individual clones for EST clusters 1 and 2 completely. Published sequences for EC3 and EC4 were identical to the genomic DNA, and thus clones were not resequenced.

DNA sequencing and sequence analysis. We isolated DNA of genomic PAC clones and subcloned it into M13 as described[17]. We carried out PAC contig construction, subcloning, sequencing and first-pass annotation using RUMMAGE-DP (ref. 17, and G.G. *et al.*, in preparation). cDNA clones and RT–PCR products were sequenced using appropriate primers and the BigDye terminator cycle sequencing reaction kit (PE Applied Biosystems), and analysed on an ABI Prism 377 DNA sequencer. We analysed the sequences with the DNASTAR sequence analysis program package.

Primer selection. All primers were selected with either the OLIGO program (version 4.1, National Biosciences) or the on-line program PRIMER3 (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi/), and were synthesized by Eurogentec.

PCR-based gene analysis. As templates for RT–PCR, we used either cDNA, which we prepared from human fetal brain poly(A)$^+$ RNA (Clontech) using the GeneAmp RNA PCR Kit (Perkin-Elmer) according to the recommendations of the supplier, or Marathon-ready cDNA from human fetal brain (Clontech). For long-range PCR, we used the Advantage cDNA PCR kit

(Clontech). RT–PCR was performed with the following amplimers: RT1, PM144 and PM129; RT2, PM145 and PM129; RT3, PM104 and PM83; RT4, PM148 and PM154; RT5, PM101 and PM61. We sequenced RT–PCR products with the PCR and internal primers to prove their identity. We prepared hybridization probes 1 and 2 by PCR with primers PM94/PM91 and PM67/PM66 using DNA of PAC24 and PAC22 as templates and subsequent radiolabelling. Primer sequences are available (http://www.uni-essen.de/~thg020/).

Northern-blot analysis. We probed commercial northern blots containing purified poly(A)$^+$-RNA (~2 μg) from four human fetal tissues or eight human adult tissues (Clontech) with the radioactively labelled probes. Hybridizations were done overnight at 42 °C. Posthybridization washes were done at a final stringency of 3×SSC, 50 °C for 10 min. Exposure was for 2 d.

cDNA library screening. We screened 1×10^6 plaques of a random and oligo(dT)-primed human fetal brain 5′-stretch cDNA library in λgt10 (Clontech) with the hybridization probes 1 and 2. We determined the insert sequences of positive clones using vector-specific primers as well as gene-specific primers.

Mutation analyses. This study was approved by the local ethics committee at the Universitätsklinikum Essen. We prepared DNA of patients and control individuals from peripheral blood lymphocytes using standard procedures. Overlapping amplimers were established for the entire coding region of *TRPS1*. Primers were selected as described above and are available (http://www.uni-essen.de/~thg020/). We performed genomic PCR in reactions (50 μl) containing genomic DNA (100 ng) under standard conditions (PE Applied Biosystems), and purified the products using Microcon YM-100 filter devices (Amicon). We sequenced both strands using the same primers as for amplification and the ABI Prism BigDye terminator cycle sequencing reaction kit (PE Applied Biosystems), and analysed them on an ABI Prism 377 DNA sequencer. We designated the mutations as recommended by the Nomenclature Working Group[18].

GenBank accession numbers. Genomic DNA sequence of PACs 22–24, AF178030; *TRPS1* cDNA, AF183810. Representative clones for each EST cluster are as follows: EC1, AA662457, IMAGE clone 1218102; EC2, H53479/H53854, IMAGE clone 202658, and R83254/H50546, IMAGE clone 186882; EC3, AA470941/AA470973, IMAGE clone 771530; EC4, R54192/R54193, IMAGE clone 41822.

Acknowledgements

We thank the patients and their clinicians; S. Groβ, M. Klutz and S. Rothe for technical assistance; U. Claussen, B. La Pillo, J. Nardmann, M. Wagner and D. Wells for collaboration during initial stages of this project; D. Lohmann for help with the artwork; and E. Passarge for continuous support. Part of this research was supported by the Deutsche Forschungsgemeinschaft and the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie.

Received 20 August; accepted 1 December 1999.

1. Giedion, A., Burdea, M., Fruchter, Z., Meloni, T. & Trosc, V. Autosomal dominant transmission of the tricho-rhino-phalangeal syndrome. Report of 4 unrelated families, review of 60 cases. *Helv. Paediatr. Acta* **28**, 249–259 (1973).
2. Lüdecke, H.-J. *et al*. Molecular dissection of a contiguous gene syndrome: localization of the genes involved in the Langer-Giedion syndrome. *Hum. Mol. Genet.* **4**, 31–36 (1995).
3. Hou, J. *et al*. A 4-megabase YAC contig that spans the Langer-Giedion syndrome region on human chromosome 8q24.1: use in refining the location of the trichorhinophalangeal syndrome and multiple exostoses genes (*TRPS1* and *EXT1*). *Genomics* **29**, 87–97 (1995).
4. Lüdecke, H.-J. *et al*. Genes and chromosomal breakpoints in the Langer-Giedion syndrome region on human chromosome 8. *Hum. Genet.* (in press).
5. Ahn, J. *et al*. Cloning of the putative tumour suppressor gene for hereditary multiple exostoses (*EXT1*). *Nature Genet.* **11**, 137–143 (1995).
6. Nardmann, J., Tranebjærg, L., Horsthemke, B. & Lüdecke, H.-J. The tricho-rhino-phalangeal syndromes: frequency and parental origin of 8q deletions. *Hum. Genet.* **99**, 638–643 (1997).
7. Kozak, M. Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**, 187–208 (1999).
8. Jarrousse, A.S., Petit, F., Kreutzer-Schmid, C., Gaedigk, R. & Schmid, H.P. Possible involvement of proteasomes (prosomes) in AUUUA-mediated mRNA decay. *J. Biol. Chem.* **274**, 5925–2930 (1999).
9. Dai, K.S. & Liew, C.C. Characterization of a novel gene encoding zinc finger domains identified from expressed sequence tags (ESTs) of a human heart cDNA database. *J. Mol. Cell. Cardiol.* **30**, 2365–2375 (1998).
10. Brown, K.E. *et al*. Association of transcriptionally silent genes with IKAROS complexes at centromeric heterochromatin. *Cell* **91**, 845–854 (1997).
11. Sun, L., Liu, A. & Georgopoulos, K. Zinc finger-mediated protein interactions modulate *Ikaros* activity, a molecular control of lymphocyte development. *EMBO J.* **15**, 5358–5369 (1996).
12. Boulikas, T. Putative nuclear localization signals (NLS) in protein transcription factors. *J. Cell. Biochem.* **55**, 32–58 (1994).
13. Pehlivan, T. *et al*. *GATA4* haploinsufficiency in patients with interstitial deletion of chromosome region 8p23.1 and congenital heart disease. *Am. J. Med. Genet.* **83**, 201–206 (1999).
14. Kalff-Suske, M. *et al*. Point mutations throughout the *GLI3* gene cause Greig cephalopolysyndactyly syndrome. *Hum. Mol. Genet.* **8**, 1769–1777 (1999).
15. Fisher, E. & Scambler, P. Human haploinsufficiency—one for sorrow, two for joy. *Nature Genet.* **7**, 5–7 (1994).
16. Ioannou, P.A. *et al*. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* **6**, 84–89 (1994).
17. Glöckner, G. *et al*. Large-scale sequencing of two regions in human chromosome 7q22: analysis of 650 kb of genomic sequence aound the *EPO* and *CUTL1* loci reveals 17 genes. *Genome Res.* **8**, 1060–1073 (1998).
18. Antonarakis, S.E. & the Nomenclature Working Group Recommendations for a nomenclature system for human gene mutations. *Hum. Mut.* **11**, 1–3 (1998).
19. Yi, Y. *et al*. The ancestral gene for transcribed, low-copy repeats in the Prader-Willi/Angelman region encodes a large protein implicated in protein trafficking, which is deficient in mice with neuromuscular and spermiogenic abnormalities. *Hum. Mol. Genet.* **8**, 533–542 (1999).