SUPPLEMENTARY MATERIALS AND METHODS

Cytogenetic, Fluorescence In Situ Hybridization (FISH) Analyses

G-banded metaphase chromosomes were analyzed at 550-band resolution. To map the breakpoints on chromosomes 8 and X by FISH, a series of bacterial artificial chromosomes (BAC) and P1 artificial chromosomes (PAC) clones were selected from the UCSC Genome Browser (http://genome.ucsc.edu, release NCBI36/hg18, March 2006) covering 8q24.3 (16 clones) and Xq24-25 (18 clones). BAC/PAC DNA was isolated using standard protocols. FISH probes were labelled, by nick translation (Roche Diagnostics, Mannheim Germany) with biotin-16-dUTP, and FISH-analyses were performed according to standard protocol.

Array-Comparative Genome Hybridisation (a-CGH)

Array-CGH was carried out on genomic DNA using a whole genome oligonucleotide microarray platform (Human Genome CGH Microarray 244A Kit; Agilent Technologies, Santa Clara, California, USA). This array consists of approximately 236,000 60-mer oligonucleotide probes with a spatial resolution of 8.9 Kb. Experiments were performed following manufacturer's instructions. Slides were scanned using a G2565BA scanner, and analyzed using Agilent CGH Analytics software ver. 4.0.81 (Agilent Technologies Inc.) with the statistical algorithm ADM-2 and a sensitivity threshold of 6.0. At least three consecutive probes had to be aberrant to be identified as significant copy-number change.

X-Inactivation Study

To determine X-inactivation pattern, a CAG triplet repeat in the first exon of the androgen receptor (*AR*) gene (Xq12) was analyzed by methylation-specific PCR as described [1]. The genomic DNA was amplified with and without prior Hae II methylation sensitive restriction enzyme digestion. Using a forward fluorescent labelled primer, we amplified the *AR* CAG-containing polymorphic region in the proband and her relatives, used as controls. Reaction products were electrophoresed on an ABI Prism 3100 Avant automatic sequencer (Applied Biosystems).

Breakpoint junction definition

To determine the boundaries of the breakpoint on chromosomes X and 8, we used a long range PCR strategy with various combinations of primers and the Expand Long Template PCR kit in the conditions specified (Roche Diagnostics). The breakpoint junction on chromosome 8 was amplified using: 10 µM primers 5'-tcttctatttagtatttctgg and 5'-tgctgagaatgatggtttcc, 144 µM dNTPs, 1.8 mM $MgCl_2$, 90 ng of genomic DNA, and 1 U of Taq Gold (Applied Biosystems). The thermal cycling

conditions were 7 min at 95°C, followed by 30 cycles 30 sec at 95°C, 30 sec at 51°C, 1 min at 72°C; a final extension of 10 min at 72°C. The breakpoint junction on chromosome X was amplified using: 10 µM primers 5'-ctgtgtctgtgtagaaagag and 5'-caggtgtggtaatgcatgcctggatc, 144 µM dNTPs, 2 mM $MgCl_2$, 40 ng of genomic DNA, and 1 U of Taq Gold (Applied Biosystems). The thermal cycling conditions were 7 min at 94°C, followed by 14 cycles 30 sec at 94°C, 30 sec at 62 - 0.5°C, 1 min at 72°C, 30 cycles 30 sec at 94°C, 30 sec at 55°C, 1 min at 72°C; a final extension of 10 min at 72°C. PCR products were purified using a ExoSAP strategy and directly sequenced by the Big-Dye cycle sequencing kit ver.1.1 and an ABI Prism 3100 Avant automatic sequencer (Applied Biosystems).

Expression analyses

Amplification of part of the *PTK2* cDNA was obtained using primers 5'-tattttgaagacttgagttatttcaga; 5'-gtgctctggtacaaagcatttctg under standard conditions. Patient was found heterozygous for the SNP rs#7460 in exon 32. TaqMan real-time quantitative PCR analysis was used to measure expression in human cells as follows: (a) *PTK2*, FAM-labeled pre-designed TaqMan gene expression assays (Hs00178587_m1, Applied Biosystems); (b) *THOC2* exons 1-2, primers 5'-caaggaatcttcagcaaagctc; 5'-ccatttctctctccttactcttgg; #68 probe (Roche Diagnostics); (c) *THOC2* exons 33-34, FAM-labelled pre-designed TaqMan gene expression assays (Hs00396154_m1, Applied Biosystem); (d) *XIAP*, FAM-labelled pre-designed TaqMan gene expression assays (Hs01597786_g1, Applied Biosystem); (e) *TBP* reference gene, VIC-labeled pre-designed TaqMan gene expression assays (Hs00427620_m1, Applied Biosystems). Reactions were carried out on an ABI 7500 Fast real-time PCR machine using the ABI TaqMan Universal PCR master mix according to the manufacturer's instructions (Applied Biosystems). Efficiencies of the assays were similar and in a range 90-110%. Patient, and three gender matched unrelated healthy controls were run in triplicate; the mean Ct value was used for calculations using the ΔΔCt method [2]. In situ hybridizations using antisense RNA in mouse brain, were obtained for *Thoc2* from "The Allen Brain Atlas" [3] (Allen Institute for Brain Science. ©2009. Available from: http://mouse.brain-map.org, experiment 69444837).

Western blot assay

Proteins from cell lysates (20 µg) were separated by precast gradient gel (4-15%) (Bio-Rad) and transferred to 0.45 µm nitrocellulose membrane (BIO-RAD). After blocking with 5% non-fat milk in Tris-buffered saline containing 0.1% Tween (TBS-T), protein blots were incubated with a specific antibody (PTK2, [4] antibody against the C-terminus of THOC2 [5] and BACT (Novus Biological) followed by incubation with a peroxidase-conjugated secondary antibody in blocking

buffer. Protein bands were visualized with the enhanced chemiluminescence method according to manufacturer's instructions (CYANAGEN).

Mice tissue collection and RNA isolation

The experimental protocol was approved by the Bioethical Committee of the University of Turin and by the Italian Ministry of Health. Mice were sacrificed under deep general anaesthesia (ketamine, 100 mg/kg; Ketavet, Bayern, Leverkusen, Germany; xylazine, 5 mg/kg; Rompun; Bayer, Milan, Italy). Brains were temporarily placed in an ice-cold artificial cerebro-spinal fluid containing 125 mM NaCl, 2.5 mM KCl, 2 mM $CaCl_2$, 1 mM $MgCl_2$, 1.25 mM NaH_2PO_4, 26 mM $NaHCO_3$, 20 mM glucose, bubbled with 95% O_2 and 5% CO_2. For the analysis on separated brains regions, cerebellums were manually dissected. All samples were rapidly frozen in 2-methylbutane and stored at -80°C. Total RNA was extracted with TRIzol Reagent (Invitrogen Life Technologies) in accordance with the manufacturer's instructions. One microgram of total RNA was reverse-transcribed using the High-Capacity cDNA Archive Kit (Applied Biosystem, Monza, Italy), accordingly to the manufacturer's instructions. cDNA samples were stored at -20°C.
TaqMan real-time quantitative PCR analysis was used to measure expression in mouse tissues using the following assays: *Thoc2*, primers 5'-gatctggccgaaataaccac; 5'-gcagttagaagccatgactg; UPL#46 human set probe (Roche Diagnostics); *Ptk2* primers 5'-gaaaagcggctgtagtcca; 5'-ccattcttttgctagatgctaggt UPL#70 human set probe (Roche Diagnostics); reference gene was *Hmbs* primers 5'-tccctgaaggatgtgcctac; 5'-cacaagggttttcccgttt; UPL#10 human set probe (Roche Diagnostics). Reactions were carried out and analysed as described above. Experiments were performed three times in triplicate.

In vitro transcription/translation

To obtain the chimeric *PTK2-THOC2* amplimer, we amplified the patient's cDNA using a forward primer on exon 1 of the *PTK2* gene (*T7 promoter* - specific primer, 5'-ggatcctaatacgactcactatagggagacc – accatggcggccgcggctgtgtgggttcccg) and a reverse primer on exon 12 of the *THOC2* gene (5'-ctccttcataaatgacttgcctatg). As control a *THOC2* partial transcript was obtained using primers in exon 1-exon 12 (*T7 promoter* - specific primer, 5'-ggatcctaatacgactcactatagggagacc - gcgtgaggcgtgggaggaagcgcgg). PCR conditions were: 200 nM of each primer, 200 µM dNTPs, 1x KAPA2G Buffer B, 100 ng of cDNA, and 1 U of KAPA2G Fast HotStart DNA Polymerase (KAPABiosystems). Transcription/translation was performed using TnT T7 Quick for PCR as described (Promega).

Caenorhabditis elegans models

C.elegans viability and development. The ability of animals to survive and develop to adulthood compared to a wild-type strain was assessed by scoring for three consecutive generations animals fed bacteria expressing empty-vector or bacteria expressing either of dsRNA against *kin-32* or *thoc-2.*

Fecundity/fertility. The total number of progeny (brood size or fecundity) and the number of eggs hatched (fertility) from three independent gravid adults (in two independent experiments) were scored in each of the above-described experimental conditions.

Sensory neurons functionality. Different types of chemosensory neurons respond to a variety of volatile and water-soluble chemicals. The functionality of a specific subset of animal sensory neurons (AWA, AWB, AWC, ASE) was assessed by quantifying their attraction or repulsion to specific chemicals (respectively: pyrazine, nonanone, benzaldheide, NH_4Acetate). Briefly, sodium azide (NaN_3) used to anesthetize worms is placed on buffered agar 180 degrees opposite on a 10 cm dish; the attractant (or repellent) is then placed on one NaN_3 spot, and ethanol (neutral odor for the worms in which attractant is diluted) on the other spot; a population of 80-100 age-synchronized animals are spotted in the centre of the testing plate and the number of worms at attractant and control are counted every 15 minutes for two hours to calculate the Chemotaxis Index (CI). CI=(A-B)/(A+B+C), where A is the number of worms at attractant, B is the number of worms at control and C is the number of animals which did not reach any of the two spots at the end of the two hours. For a population of young (3 days old), wild-type animals, a good CI is around 0,8 for attractants and –0.8 for repellents after two hours (CI=0 means no attraction, while CI=1 or –1 represent maximum attraction or repulsions, but there are always some animals that, also for a wild-type strain, remain randomly dispersed in the assay plate or reach the control spot instead of the attractant). *Che-3* sensory defective mutant was used as positive control.

Motoneuron activity. We calculated animal spontaneous locomotion counting the number of body bends (i.e., changes in the body bend at the mid-body point) per minute on solid agar plates with no bacteria. One bend was counted every time the mid-body reaches a maximum bend in the opposite direction from the bend last counted. Body bend was checked in at least 15 animals for 15 seconds.

Proprioception. Gentle touch to the body is sensed by touch receptor neurons, which extend their neurites in close contact with the cuticle along their entire length. In this assay animals were gently touched with an eyelash ten times, alternating six head (backward movement) and six tail (forward movement) touches, and a score was given for the number of positive responses. No more than six touches per side were done to avoid habituation.

All behavioral assays were performed on young animals on their first fertile day (3 day-old animals).

Plasmid preparation, transfections and silencing quantification

pGIPZ Lentiviral shRNAmir clones (Thermo Scientific, Open Biosystem) were grown for 18-19 hours at 37°C in LB broth media plus 100 μg/ml ampicillin. Plasmid DNA was extracted using Promega Pure Yield ™ Plasmid Miniprep System. We seeded $4 \cdot 10^4$ HeLa cells per well in 24-well plates and transfected them 24 h later with 500 ng of plasmid per well. Each plasmid was delivered using FuGene HD Transfection reagent (Roche) according to the manufacturer's protocol. Puromycin selection was done after 72 h using a puromycin working concentration of 3 μg/ml for 12 hours. Transfection efficiency was determined counting the percentage of fluorescent cells by flow cytometry using GFP-expressing plasmid . To analyze messenger RNA knockdown, cell were lysed and total RNA collected using the Cells-to-Ct kit (Applied Biosystem) in accordance with the manufacturer's protocol.

THOC2 mRNA level was determined by real-time RT-PCR as described above.

Mutation screening

Forty-two sporadic patients (33 males and 9 females) with mental retardation and childhood onset cerebellar ataxia were collected for mutation screening. PCR amplification and DHPLC analysis were performed as follow: *PTK2* and *THOC2* gene-coding regions were divided into 32 and 38 amplimers, respectively (sequences available upon request). Amplicons were run on DHPLC (Transgenomic WAVE System) using melting temperatures determined by the DHPLC Melt software (available upon request) [6]. A normal control profile was always compared with that from patient. For *THOC2*, two male patients were mixed to allow the formation of heteroduplexes. PCR products showing a DHPLC peak shift were purified using the ExoSAP system (Amersham-Pharmacia Biotech, Freiburg, Germany), and sequenced by the BigDye cycle sequencing kit and an ABI Prism 3100 Avant automatic sequencer (Applera). To determine whether a nucleotide change could affect splicing, we scored these substitutions by two web available softwares: the MaxENT (http://genes.mit.edu/burgelab/maxent//Xmaxentscan_scoreseq_acc.html)[7], and the Splice Site Prediction by Neural Network (http://www.fruitfly.org/seq_tools/splice.html) [8].

Whole gene deletions were screened using a real-time PCR assay centred on exons 3 for *PTK2* (conditions available upon request).

SUPPLEMENTARY RESULTS

Mapping of the translocation breakpoints by FISH

We performed FISH analysis on patient's metaphase spreads to characterize the breakpoint on both chromosome X and 8. We selected Bacterial Artificial Chromosomes (BACs) and P1 Artificial Chromosomes (PACs) . FISH analyses and clones that allowed the breakpoints identifications are reported in Fig. 1D and scheme 1E. On chromosome Xq24-25, we identified one PAC clone (RP5-931E15) hybridising both derivatives (Fig. 1D, 1E) and two flanking clones hybridising to chromosome X-der(X) (RP5-506G2), and chromosome X –der(8) (RP11-325K14). On chromosome 8q24 the breakpoint could be located in a region of ~ 33 kb overlapping two BAC clones (RP11-691F18 and RP11-159E16, Fig. 1D and E) which hybridised both derivatives (Fig. 1D).

Mutation analysis

With the aim of identifying a second patient with a mutation in *PTK2* or *THOC2* genes, we selected 42 pediatric cases (33 males and 9 females) with mental retardation and ataxia negative for mutations in known genes. The *PTK2* gene was screened in 18 patients (9 males and 9 females) by real time PCR for whole-gene deletions / duplications and by dHPLC for point mutations (see Table 1 in the supplement). All 42 cases were screened for *THOC2* point mutations by dHPLC (see Table 2 in the supplement). No mutation was detected in either screening.

SUPPLEMENT Figure1

Supplement figure 1. Patient's MRI at 15 months and 6 yr

T1-weighted magnetic resonance images of the patient at 15 months and 6 yr. Sagittal, axial and coronal sections showed cerebellar hypoplasia of the hemispheres and vermis with enlargement of the IV ventricle and cisterna magna.

Supplement figure 2. FISH analysis and sequences alignment at the two breakpoints

A) Schematic representation of the Xq25 and 8q24.3 regions involved in the translocation. Genes and their transcriptional orientation are indicated as dark grey arrows. FISH probes used to define the breakpoint interval are indicated as grey bars. The breakpoint within *PTK2* and *THOC2* gene are enlarged. The repetitive elements MER4/AluJ on chromosome 8 and SVA on chromosome X were involved. B) Sequences alignment at the two breakpoints showed the involvement of nonhomologous regions and a 88 bp deletion at chromosome X.

REFERENCES

1 Kubota T, Nonoyama S, Tonoki H*, et al.* A new assay for the analysis of X-chromosome inactivation based on methylation-specific PCR. *Hum Genet* 1999;**104**(1):49-55.

2 Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001;**25**(4):402-8.

3 Lein ES, Hawrylycz MJ, Ao N*, et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 2007;**445**(7124):168-76.

4 Retta SF, Barry ST, Critchley DR*, et al.* Focal adhesion and stress fiber formation is regulated by tyrosine phosphatase activity. *Exp Cell Res* 1996;**229**(2):307-17.

5 Masuda S, Das R, Cheng H*, et al.* Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev* 2005;**19**(13):1512-7.

6 Jones AC, Austin J, Hansen N*, et al.* Optimal temperature selection for mutation detection by denaturing HPLC and comparison to single-stranded conformation polymorphism and heteroduplex analysis. *Clin Chem* 1999;**45**(8 Pt 1):1133-40.

7 Eng L, Coutinho G, Nahas S*, et al.* Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: maximum entropy estimates of splice junction strengths. *Hum Mutat* 2004;**23**(1):67-76.

8 Reese MG, Eeckman FH, Kulp D*, et al.* Improved splice site detection in Genie. *J Comput Biol* 1997;**4**(3):311-23.

9 Goslin K, Banker G. Rapid changes in the distribution of GAP-43 correlate with the expression of neuronal polarity during normal development and under experimental conditions. *J Cell Biol* 1990;**110**(4):1319-31.

10 Ying QL, Smith AG. Defined conditions for neural commitment and differentiation. *Methods Enzymol* 2003;**365**:327-41.