ONLINE METHODS

Patients

The two affected brothers were born to nonconsanguineous parents with no family history of neurodevelopment disorders. Both patients were negative for FRA(X) syndrome and chromosome microarray tests. The parents and the three female siblings are healthy.

Patient 1 was born at term after caesarian section weighing 4,160 g. Examined at the age of 34, presented at physical examination: height 1.81 m, weight 86 kg, hair whorls, thick eyebrows (without synophris), large mouth, widely spaced nipples, normal genitalia and shortened Achilles tendons. Temporal irritative focus was detected by routine EEG at the age of 6, but without any convulsive or other epilepsy manifestations. He presented mild intellectual disability, without significant delayed motor milestones, but exhibited stuttering and difficulties in the articulation of words. He attends to a special education school, having achieved a level equivalent to 3rd-4th degree of elementary schooling.

Patient 2 was born at term after caesarian section weighing 3,890 g. Physical examination at the age of 32 showed: height 1.71 m, weight 55 kg, moderate degree of frontal balding, hair whorls, thick eyebrows (without synophris), large mouth, normal genitalia. Similarly to his brother, he presented a mild intellectual disability; he did not present significant delayed motor milestones, with pronounced difficulties in the

articulation of words. He also attends to a special education school, having achieved a level equivalent to 3rd-4th degree of elementary schooling.

This study was approved by the ethics committee of the Institute of Biosciences (University of São Paulo, São Paulo, Brazil), and written informed consent was obtained from parents. The current work is in compliance with all relevant ethical regulations.

Whole Exome Sequencing

DNA samples were prepared using the AmpliSeq Exome library, according to the manufacturer's specifications. The AmpliSeq libraries were single-end sequenced on an Ion ProtonTM System at Beijing Genomics Institute (BGI, Beijing, China). The raw reads were aligned to the reference genome (GRCh37/hg19), using BWA[51], and pre-processed according to GATK toolkit[52]. Filtering and priorization of variants were conducted using VarSeq® software (Golden Helix, Bozeman, MT, USA). The variants were filtered per population frequency (<0.01), quality (phred quality ≥ 20, genotype quality ≥ 20), read depth (≥ 10).

Sanger Sequencing

Sanger sequencing was used to confirm the presence of the variants considered potentially pathogenic and segregation studies (primer sequences available upon request).

X-inactivation analysis

X-inactivation tested by determining the methylation status of the androgen receptor gene (AR) was evaluated as described previously[53].

Cloning and mutagenesis

The wild-type UBE2A gene and Q93E mutant were amplified by PCR from mother´s patients and index cDNA, respectively, and cloned into pET28a vector (Novagen), using NdeI and XhoI restriction sites. Human PCNA gene was also inserted into pET28a vector using the same restriction sites. Human RAD18 was cloned into the restriction sites BamHI e XhoI of a modified version of pETSUMO (Invitrogen). Expression plasmids for human Uba1 and Ub (both in pET3a) were donated by Drs. Titia Sixma and Chris Hill, Addgene plasmids #63571 and #61937, respectively. All point mutations in UBE2A (Q93E, Q93A, Q93R, C88S, C88S/Q93E, R7W, R11Q, G23R) were created using basic PCR-based site-directed mutagenesis and confirmed by Sanger sequencing.

Protein expression and purification

Wild-type UBE2A (residues 1-152) and mutants (Q93E, Q93A, Q93R, C88S, C88S/Q93E, R7W, R11Q, G23R) were expressed from pET28a (Novagen) as an N-terminal His6-tag fusion. Rosetta II E. coli cells (Novagen) were used to express the proteins in LB media at 37°C for 5 hours at 200 rpm. The culture temperature was turned down to 18 °C for 1 hour before induction with 0.2 mM IPTG (isopropyl β -D-1-thiogalactopyranoside)

and cells were grown at 18°C for 16 h, then harvested by centrifugation (5,000 × g for 10 min at 4 °C). Cells were disrupted by sonication in lysis buffer (50 mM Tris pH 7.5, 300 mM NaCl, 10% glycerol, 2 mM β-mercaptoethanol and 1 mM phenylmethylsulfonyl fluoride) containing 0.2 mg/mL lysozyme and centrifuged to remove cellular debris (40 000 x g for 60 min at 4 °C). Proteins were first purified on a HisTrap column (GE Healthcare) pre-equilibrated with binding buffer containing 50 mM Tris pH 7.5, 300 mM NaCl, 2 mM β-mercaptoethanol and eluted using a linear gradient of same buffer containing 500 mM imidazole, in an AKTA Purifier system (GE Healthcare). The His6-tag was cleaved by thrombin protease and the proteins were further purified by a Superdex 75 column (GE Healthcare) with buffer containing 50 mM Tris pH 7.5, 300 mM NaCl, 1 mM dithiothreitol. After concentration to 10 mg/ml, proteins were flash-frozen in liquid nitrogen and stored at −80°C. Recombinant purified UBE2A proteins are shown in Supplementary Figure 12. Human Uba1, Ub, SUMO-RAD18 and PCNA were expressed and purified as previously described[21,54,55].

Crystal Structure Determination

Crystals were obtained by sitting-drop vapor diffusion at 18°C mixing equal volumes of protein at 10 mg/mL and the reservoir solution containing 50 mM di-Sodium succinate pH 7.0 and 12-14% PEG 3350. Harvested crystals were cryoprotected with 10% ethylene glycol added to the mother liquor (v/v). Data were collected at 1.46 Å at the beamline W01B-MX2 (LNLS) at 100 K, processed using XDS[56] and merged and scaled with Aimless[57]. The structures were solved by molecular replacement using MolRep[57].

The human UBE2B protein (PDB ID code 2YB6)[21] was used as search model for WT UBE2A, which structure was used as search model for Q93E mutant. The models were refined using Phenix program (phenix.refine)[58] and the quality of the final models was assessed with MolProbity. The crystallographic parameters and final refinement statistics are summarized in Supplementary Table 1. Both final models contained 97.3% of residues in favored regions and 2.7% in allowed regions of the Ramachandran plot (no outliers).

NMR spectroscopy

^{15}N-labeled WT- and Q93E-UBE2A proteins were produced in M9 minimal media supplemented with 1 g/L $^{15}NH_4Cl$ (Cambridge Isotope Laboratories). To produce $^{13}C/^{15}N$-labeled Q93E-UBE2A the M9 was also supplemented with 4 g/L [$^{13}C_6$]-D(+)-glucose (Cambridge Isotope Laboratories). The expression and purification protocols were performed as described for the unlabeled proteins. The gel filtration final buffer contained 50 mM K_2HPO_4-KH_2PO_4 (pH 8.0), 200 mM NaCl, 1 mM DTT. Labeled proteins were concentrated to 750 μM and 10% D_2O was added. NMR spectra were recorded at 25^{0}C on an Agilent Inova 500-MHz spectrometer. The spectra were processed with NMRPipe/NMRDraw[59] and analyzed with NMRView[60]. The backbone assignment of Q93E-UBE2A was determined using the three-dimensional experiments HNCACB, CBCA(CO)NH, HNCO, and HN(CA)CO and the graphical interface Smartnotebook.

NMR Thioester and Oxyester Reactions

Thioester reaction was performed in the NMR tube (in 200 µL) using 60 µM ^{15}N-UBE2A, 180 µM ubiquitin and 3 µM Uba1 in buffer containing 50 mM Tris pH 8.0, 150 mm NaCl, 5 mM MgCl, 0.5 mM DTT. For oxyester reactions, 400 µM ^{15}N-C88S-UBE2A or ^{15}N-C88S/Q93E-UBE2A was mixed with 650 µM ubiquitin and 6 µM Uba1, or conversely, 200 µM ^{15}N-Ub was put together with 325 µM C88S-UBE2A or C88S/Q93E-UBE2A and 3 µM Uba1, in the same buffer used for thioester reaction. All reactions were conducted at 30^{0}C and started with addition of 10 mM ATP. Sequential (^{1}H,^{15}N)-HSQC spectra were acquired on a 600-MHz spectrometer (Agilent Inova) equipped with a cryogenic probe. We used the backbone resonance assignment of apo ubiquitin available on BMRB (code 4769).

***In vitro* Ubiquitination Assay**

Ubiquitination assays were performed at 32^{0}C using 1 µM human recombinant Uba1, 30 µM human recombinant ubiquitin and 20 µM WT- or Q93E-UBE2A in the ubiquitination buffer: 50 mM Tris pH 8.0, 50 mM NaCl, 50 mM KCl, 10 mM MgCl2, 0.1 mM DTT. Reactions were started by addition of 3 mM ATP and quenched with an equal volume of 2X SDS-PAGE buffer (reducing or non-reducing). Samples collected in specific time points were resolved by SDS-PAGE followed by Western blot using anti-ubiquitin antibody (P4D1 - Santa Cruz). In vitro ubiquitination assay in different pHs were conducted identically to described above replacing the buffer: sodium acetate for pH 5, MES for pH 6, Tris for pH 7-9 and Caps for pH 10-11. For intrinsic reactivity assays, WT- and Q93E-UBE2A were charged for 15 minutes at 32°C in ubiquitination buffer before

being mixed with 10 mM EDTA and 25 mM of lysine or 5 mM hydroxylamine, and incubated at 32°C. Samples were collected at several time points during the reaction, quenched in non-reducing SDS-PAGE buffer and visualized by Coomassie stained SDS-PAGE. PCNA mono-ubiquitination was performed by mixing 1 μM Uba1, 30 μM ubiquitin, 20 μM WT- or Q93E-UBE2A, 5 μM SUMO-Rad18 and 1 μM PCNA in the ubiquitination buffer supplemented with 2 μM $ZnCl_2$ and incubated at 32^0C in the presence of 3 mM ATP, followed by anti-PCNA (P8825 - Sigma) Western blot.

Single-turnover diubiquitin formation assay

Diubiquitin formation assay was conducted in a two-step reaction. First, charging of UBE2A was achieved by mixing 20 μM of WT- or Q93E-UBE2A with 1 μM E1, 30 μM lysine-free ubiquitin (K0-Ub) and 3 mM ATP in the ubiquitination buffer (pH 8). After incubation for 15 minutes at 37 °C, reaction was stopped by addition of 25 mM EDTA and then diluted four times in ubiquitination buffer at pH 8 or 9 (both using 50 mM Tris-HCl), resulting in final concentrations of 0,25 μM E1, 5 μM UBE2A and 7,5 μM lysine-free ubiquitin. Di-ubiquitin formation was triggered by addition of 7,5 μM WT-ubiquitin. Reactions were kept at 32°C, samples were collect at different timepoints and immediately quenched in reducing SDS-PAGE buffer. Results were evaluated by western blot using anti-ubiquitin antibody (P4D1 - Santa Cruz), followed by quantification of corresponding di-ubiquitin band with ImageJ.

Molecular Modeling

UBE2A was covalently bound to ubiquitin using as template the UBE2N~Ub structure (PDB ID code 5aiu)[30]. UBE2A was superposed to UBE2N and the complex formed by UBE2A and ubiquitin was submitted to energy minimization using steepest descent followed by simulated annealing protocol (SD/SA) to remove bad contacts, fixing backbone atoms of UBE2A and ubiquitin. Following, we created the thioester bond between the carboxyl group of the terminal glycine residue of ubiquitin and the sulfhydryl group of catalytic Cys88 of UBE2A. To correct the bond angle and length, a new round of energy minimization using SD/SA was performed, fixing the backbone atoms of Cys88. All steps were performed using YASARA software and the AMBER03 force field. To position a second free-ubiquitin molecule with a lysine positioned to attack the thioester bond, we used the template of Mms2-Ubc13 covalently bound to ubiquitin and presenting the Lys63 residue from another symmetric lysine positioned close to the thioester bond (PDB ID code 2gmi)[37], when applied the symmetry operators of $P2_12_12_1$ space group using Pymol. We superposed UBE2A to Ubc13 to position the second ubiquitin generated by symmetry close to thioester bond. Then, the backbone atoms of UBE2A~Ub was fixated and an energy minimization using SD/SA was carried out to eliminate bad contacts.

Statistical analysis

Crystallographic statistics are indicated in Supplementary Table 1. All western blot experiments were repeated independently three times with similar results. (^{1}H,^{15}N)-HSQC spectra of UBE2A proteins (WT, Q93E, C88S, C88S/Q93E) were

repeated independently at least two times with similar results. Each oxyester and thioester reaction and subsequent NMR data collection was performed once. Data from discharge assays, diubiquitin formation and PCNA monoubiquitination assays were derived from three independent experiments and are presented as the mean ± standard deviation. The diubiquitin production rates were determined from a linear regression. The PCNA monoubiquitination formation data were fitted according to the equation: PCNA-Ub = Plateau$(1\text{-}e^{-Kt})$, where PCNA-Ub is the fraction of monoubiquitinated PCNA, Plateau is the PCNA-Ub value at infinite times, K is the rate constant, t is time, and $t_{1/2} = \ln(2)/K$.

DATA AVAILABILITY

The atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 6CYO (WT-UBE2A) and 6CYR (Q93-UBE2A).

METHODS-ONLY REFERENCES

51. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

52. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–498 (2011).

53. Allen, R. C., Zoghbi, H. Y., Moseley, a B., Rosenblatt, H. M. & Belmont, J. W. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am. J. Hum. Genet.* **51,** 1229–39 (1992).

54. Pickart, C. M. & Raasi, S. Controlled Synthesis of Polyubiquitin Chains. in *Methods in Enzymology* **399,** 21–36 (2005).

55. Zheleva, D. I. *et al.* A quantitative study of the in vitro binding of the C-terminal domain of p21 to PCNA: Affinity, stoichiometry, and thermodynamics. *Biochemistry* **39,** 7388–7397 (2000).

56. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 125–132 (2010).

57. Winn, M. D. *et al.* Overview of the CCP 4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 235–242 (2011).

58. Adams, P. D. *et al.* PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).

59. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–93 (1995).

60. Johnson, B. A. & Blevins, R. A. NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4,** 603–614 (1994).