nature research

Corresponding author(s): Dr. Andrea Cortese

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	ABI 3730xl genetic analyser (Applied Biosystems, Foster City, CA, USA), LEICA SCN400F scanner (LEICA Milton Keynes, UK), LEICA Slidepath (LEICA Milton Keynes, UK), LSM710 confocal microscope (Zeiss, Oberkochen, Germany), Odyssey Fc infrared scanner (Li-COR Bioscience), Odyssey Fc infrared scanner (Li-COR Bioscience, Lincoln, NE, USA), QuantiStudio 7 Flex (Applied Biosystems, Foster City, CA, USA).
Data analysis	NovoSort , using HTSeq, DESeq2, DEXSeq, g:Profiler; GenomeStudio, PLINK 1.9, MERLIN, Burrows-Wheeler Aligner, GATK UnifiedGenotyper , ANNOVAR, LUMPY , Integrated Genomic Viewer, GeneMapper, RBPmap, Homozygosity mapper.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The genotyping microarray data and sequence data obtained by whole-genome sequencing (A.C.; H.H.). They are not publicly available because some of the study participants did not give full consent for releasing data publicly. Since whole-genome sequence data are protected by the Personal Information Protection Law, availability of these data is under the regulation by the institutional review board. The data obtained RNA sequencing will be available on SRA upon publication.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All available samples from patients with CANVAS (29 individuals, 23 affected and 6 unaffected, from 11 families) and late onset ataxia referred to the National Hospital for Neurology and Neurosurgery in London (150 cases) were included in the study.
Data exclusions	no data were excluded from the analysis
Replication	Experiments were repeated by two experimentalists and analyzed independently leading to the same conclusions.
Randomization	Randomization is not relevant to this study. Both patients with CANVAS or late onset ataxia and healthy controls were tested for the presence of the AAGGG repeat expansion
Blinding	Blinding was not relevant to this study as no group allocation was performed in this study.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	No unique material was used except genomic DNA and tissue samples from participants. Availability of these materials are under the regulation by the institutional review board.

Antibodies

Antibodies used	1. anti- human RFC1 antibody (GeneTex , GTX129291, 1:1000) 2. anti -human p62 antibody (Abcam, ab56416, 1:500)

3. anti- human DP43 (Novus Biologicals, 2E2-D3, 1:500)
4. anti- human α-synuclein (Abcam, 4D6, 1:1000
5. anti-human phospho-Tau (AT-8, Innogenetics, 1:100)
6. anti- human βA4 (DAKO 6F3D, 1:50)
7. anti- human SMI31 antibody (Sternberger, 1:5000)
8 . anti- human γH2AX (Abcam, 1:1000)
9. anti- human β-Actin (Sigma-Aldrich, 1:2000).

Validation

As is stated on the manufacturers' website, each primary antibody has been individually validated to react against human protein. Validation for all antibodies used for immunohistochemistry has been performed in ISO15189 accredited laboratory.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

SH-SY5Y (ATCC, CRL-2266)

Authentication

SH-SY5Y were authenticated by the supplier.

Mycoplasma contamination

SH-SY5Y tested negative for mycoplasma contamination

Commonly misidentified lines
(See ICLAC register)

Not applicable

Human research participants

Policy information about studies involving human research participants

Population characteristics

All participants are Europeans. Patients are diagnosed with CANVAS or late-onset ataxia. Controls were 304 subject of European ancestry. Gender ratio was ~1:1 both for patients and controls and all were aged 35 or above. No other

Recruitment

29 individuals, 23 affected and 6 unaffected, from 11 families with a clinical diagnosis of CANVAS across four Centres: National Hospital for Neurology and Neurosurgery (London, UK), C. Mondino National Neurological Institute (Pavia, Italy), C. Besta Neurological Institute and Department of Neurology, School of Medicine (Ribeirão Preto, Brazil).
An additional 150 patients with sporadic CANVAS or late-onset ataxia (onset after 35 years of age) were identified from the neurogenetic database of the National Hospital for Neurology and Neurosurgery (London, UK). All available samples were tested so that there is no selection bias
304 controls were available from the National Hospital for Neurology and Neurosurgery (London, UK). Controls aged over 35 were chosen at random. Gender ratio was was ~1:1 for both patients and controls