

## 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

**Drosophila:** For standard RNA-seq, three independent biologically independent replicates were sequenced, each consisting of 500 ng of total RNA from six pooled Drosophila heads (18 heads total). Extracted RNA was used for library preparation according to the KAPA Stranded RNA-seq Kit with RiboErase (HMR) sample preparation guide. After quantification by Qubit and Bioanalysis, libraries were pooled for cBot amplification and sequenced on the Illumina HiSeq 3000 platform with 100 base pair paired-end sequencing. For small RNA-seq, four independent biological replicates were sequenced. 500 ng of total RNA from pooled Drosophila heads was used for library preparation according to the NEBNext small RNA sample preparation guide. Due to the abundance of the 2S rRNA in Drosophila, we included an additional 2S block step using the oligo (5'-TAC AAC CCT CAA CCA TAT GTA GTC CAA GCA/3SpC3/-3') as described<sup>48</sup>. 2S rRNA blocking was performed directly after 3' SR adapter ligation. 1  $\mu$ M of the 2S rRNA block oligo was added directly to each ligation reaction on ice, and reactions were incubated at 90°C for 30 sec, then 65°C for 5 min. After 5 min, 1  $\mu$ l of SR RT primer was added and we proceeded as described in the NEBNext protocol. After small RNA-seq libraries were subjected to quantification by Qubit and Bioanalysis, samples were pooled for cBot amplification and sequenced on the Illumina HiSeq 3000 platform with 50 base pair single-read sequencing.

**Human:** The Mayo RNA-seq Study on Neuropathological Diseases generated whole transcriptome data for cerebellum and temporal cortex samples from 312 North American Caucasian subjects. These subjects were diagnosed with Alzheimer's disease, progressive supranuclear palsy, pathologic aging or were elderly controls without neurodegenerative disorders<sup>59</sup>. We downloaded the corresponding clinical data (covariates) from the Accelerating Medicines Partnership – Alzheimer's Disease (AMP-AD) Knowledge Portal<sup>60</sup>. Specifically, we used the synapseClient R package v1.15-0 (<http://www.sagebase.org>) to download temporal cortex (syn5223705) and cerebellum (syn3817650) clinical data. Dr. Nilufer Taner at the Mayo Clinic provided us with unprocessed RNA-seq data.

## Data analysis

**Drosophila:** After standard and small-RNA sequencing, CASAVA was used for demultiplexing and fastq files were generated for each sample.

For data cleaning, SortMeRNA49 v2.1 was used to identify and exclude ribosomal RNA reads, and Trimmomatic50 v0.36 was used to remove Illumina adaptors. FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) v0.11.5 was used for a quality check before and after the above cleaning steps. The quality of bases in the cleaned reads was above 28 (Sanger/Illumina 1.9 encoding). Reads were aligned to the transposon reference FASTA files (FlyBase51 Release 6.12), and the quantity of each transposable element was calculated using Salmon52 v0.7.2. On average, 9 million reads were mapped per sample. DESeq253 v1.14.1 was used to identify sequences that were differentially expressed in tau transgenic Drosophila compared to controls. Ensembl BioMarts54 was used to assign genomic locations of differentially expressed transcripts (Supplementary Table 2). Pheatmap (<https://CRAN.R-project.org/package=pheatmap>) v1.0.8 and Pimgene55 v1.3.4 R56 packages were used to generate heatmaps. Values in transposable element heatmaps are standardized Transcripts Per Million57 (TPM). For presentation clarity in Fig. 1, we subtracted the TPM value of each transposable element from its average across all samples, and then divided the difference by the standard deviation of the TPM value of that particular transposable element. Unscaled heatmaps (Supplementary Fig. 1) represent raw TPM.

piRNA small RNA-seq analyses were similar to the above, with some exceptions. As some tRNAs are mis-annotated as piRNAs, piRNAs that were a subsequence of a tRNA (FlyBase51 Release 6.16) were removed from analysis. Reads were mapped to the remaining piRNA sequences (piRNABank58) as the reference. On average, 16 million reads were mapped per sample. piRNAs with low coverage were excluded as follows: If the sum of reads that mapped to a piRNA in all four tau transgenic Drosophila samples was less than three, we considered its expression “undetectable” in tau transgenic samples. Similarly, if the sum of reads that mapped to a piRNA in all four control normal samples was less than three, we considered its expression undetectable in controls. If a piRNA had undetectable expression values both tau transgenic and control samples, it was excluded from our analysis. Raw data for small RNA-seq are included in Supplementary Table 5, and genomic locations of differentially expressed piRNAs are included in Supplementary Table 6.

**Human:** We converted raw bam files to fastq files using the Picard SamToFastq v2.10.10 tool (<http://broadinstitute.github.io/picard>). We downloaded human transposable elements from the Genetic Information Research Institute (GIRI) RepBase64 database in fasta format (<http://www.girinst.org/repbase/update/browse.php?type=All&format=FASTA&autonomous=on&nonautonomous=on&simple=on&division=Homo+sapiens&letter=A>). The file contains 1073 unique sequences including the 549 ancestral repeats that are shared among all mammals. Using the same pipeline that we described for the analysis of Drosophila RNA-seq data, we cleaned the fastq files, aligned them to the human transposable element sequences, and performed differential expression analysis.

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

Raw counts from RNA sequencing are provided as Supplementary Tables. Full access to bam files that include Drosophila transposable element and piRNA reads will be made publicly available upon publication. Custom codes that were created for cleaning and analyzing sequencing data will be made available 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](https://www.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

Sample size was based on power analysis for the original submission. Sample sizes for TUNEL and PCNA were increased for the final submission based on a reviewer request.

### Data exclusions

Exclusion criteria were pre-established prior to data analysis.

For Drosophila piRNA small RNA-seq analyses, as some tRNAs are mis-annotated as piRNAs, piRNAs that were a subsequence of a tRNA (FlyBase51 Release 6.16) were removed from analysis. Reads were mapped to the remaining piRNA sequences (piRNABank58) as the reference. piRNAs with low coverage were excluded as follows: If the sum of reads that mapped to a piRNA in all four tau transgenic Drosophila samples was less than three, we considered its expression “undetectable” in tau transgenic samples. Similarly, if the sum of reads that mapped to a piRNA in all four control normal samples was less than three, we considered its expression undetectable in controls. If a piRNA had undetectable expression values both tau transgenic and control samples, it was excluded from our analysis.

For human RNA-seq data, 12 cortex (syn6126114) and 10 cerebellum (syn6126119) samples were excluded due to low quality by the Mayo Clinic61,62. In this study, we also excluded 9 cortex and 15 cerebellum samples that had an RNA Integrity Number (RIN)63 less than 7.

After this filtering, 80 Alzheimer's disease, 82 progressive supranuclear palsy, and 21 control cortex samples were available, as well as 76 Alzheimer's disease, 78 progressive supranuclear palsy, and 25 control cerebellum samples.

Replication	All attempts at replication were reproducible. For representative images, quantification of replicates is provided. For many experiments, multiple investigators repeated the same experiment. The number of independent experiments is presented in the figure legends and Supplementary Table 9.
Randomization	The order of Drosophila genotypes used for dissection/fixation/head homogenization was randomized for each experiment. The order of genotypes deposited in NanoString cartridges was also randomized.
Blinding	Blinding was performed for immunofluorescence and immunohistochemistry (including TUNEL and PCNA counting).

## Reporting for specific materials, systems and methods

### Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	Presented in Supplementary Table 8 Actin (Dm), Abcam #ab8227, dilution 1:50,000, Secondary rHRP, SouthernBiotech #3010-05, dilution 1:20,000; GFP (Dm), UCDavis/NIH NeuroMab #75-131, dilution 1:1000, secondary mHRP, SouthernBiotech #1010-05, dilution 1:10,000; PCNA (Dm), Dako #M0879, dilution 1:200, secondary rBio, SouthernBiotech #4010-08, dilution 1:200; piwi (Dm), Abcam #5207, dilution for WB 1:500, for IF 1:50, secondary rHRP for WB, SouthernBiotech #3010-05, dilution 1:1000, secondary r555 for IF, Thermo Fisher #A21434, dilution 1:200; cTau (Dm), Dako #A0024, dilution 1:2000,000, secondary rHRP, SouthernBiotech #3010-05, dilution 1:20,000.
Validation	GFP validated by western blot using Drosophila lines with and without transgenic GFP (data not shown); PCNA validated previously by genetic manipulation of cell cycle machinery (Khurana, V. et al. TOR-mediated cell-cycle activation causes neurodegeneration in a Drosophila tauopathy model. Current biology : CB 16, 230-241, doi:10.1016/j.cub.2005.12.042 (2006).) piwi validated by western blot, piwi RNAi Drosophila lines (Supp. Fig. 4a); cTau (Dm), Dako #A0024 - validated by western blot, tau transgenic and non-transgenic Drosophila (Supp. Fig. 2a) Actin was not validated but is widely used

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Drosophila melanogaster - equal numbers of males/females were used unless otherwise noted. Age is day 10 of adulthood unless otherwise noted.
Wild animals	No wild animals were used in this study.
Field-collected samples	No field-collected samples were used in this study.

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	<i>Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural &amp; social sciences study design questions and have nothing to add here, write "See above."</i>
Recruitment	<i>Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.</i>