Appendix

Supplementary Information

SI results

ATXN3 associates with RNAP II and key C-NHEJ proteins

To investigate whether the association of ATXN3 with C-NHEJ proteins enhanced upon DNA damage induction, we performed ATXN3 co-IP from SH-SY5Y neuroblastoma cells (either mock-treated or treated with radiomimetic Bleomycin; Bleo; a DSB inducing agent) and compared the association of C-NHEJ factors with ATXN3. We indeed observed enhanced association of ATXN3 (~1.5-2 fold) with the core C-NHEJ factors and RNAP II upon damage induction (*SI Appendix*, Fig. S1*A*), a result which is consistent with our earlier observation (1). To test whether ATXN3 associates with RNAP II and C-NHEJ proteins in a large multiprotein complex, we size fractionated HeLa-S3 cell nuclear extracts in a Sephacryl S-300 gel filtration column. Immunoblotting of eluates revealed that a significant fraction of the ATXN3 was present in a large megadalton complex, which also contained the core C-NHEJ proteins Lig IV, PNKP as well as RNAP II (*SI Appendix*, Fig. S1*B*).

SI Materials and Methods

Cell culture and various treatments

The human neuroblastoma (SH-SY5Y; ATCC CRL-2266), Human Embryonic Kidney 293 (HEK293; ATCC CRL-1573) and human bronchial epithelial BEAS-2B (ATCC CRL-9606) cells were grown at 37°C and 5% CO_2 in DMEM:F-12 (1:1, Cellgro) medium containing 10% fetal bovine serum (Sigma), 100 units/ml penicillin, and 100 units/ml streptomycin (Cellgro). To generate neural stem cells (NSCs), a normal iPSC line KYOU-DXR0109B (201B7, ATCC) was

www.pnas.org/cgi/doi/10.1073/pnas.1917280117

grown in Cell Matrix basement membrane gel (Gibco) and Pluripotent Stem Cell (PSC) SFM XF/FF media (Gibco) at 37°C and 5% CO_2. Derivation of NSCs was done using PSC neural induction medium (Gibco) as described previously (2, 3). Briefly, Essential 8 media (E8M; Gibco; required for feeder-free iPSC culture) was replaced with neural induction medium approximately 24 h after passaging iPSCs, which were then maintained in this medium for 7 days. On day 7 of neural induction, NSCs (Passage 0=P0) were cultured into Geltrex (Gibco)-coated 6-well plates and expanded in StemPRO neural stem cell SFM media. Neural induction efficiency was tested at P3 by immunofluorescence staining with a pluripotent marker (Oct4) and neural lineage stem cell marker (Nestin) as described (2).

For DNA repair studies, co-IP and DNA/RNA-ChIP, cells were either mock-treated, treated with glucose oxidase (GO, 200 ng/ml for 30 min; Sigma, Cat # G7141) or with Bleomycin (Bleo, 30 μg/ml for 1 h; Teva Chemicals USA, Cat # NDC 0703-3155-01) in reduced serum media (OptiMEM, Gibco). For heat shock-induced DNA ChIP experiments, cells were exposed to 42°C for 15 min, then either mock-treated or treated with Bleo and kept for another 45 min at the same temperature. For transcriptional inhibition studies, the cells were treated with the RNAP II inhibitor DRB (5, 6-dichloro-1-β-D ribofuranosylbenzimidazole, Sigma, 100 μM final concentration) in DMEM:F12 (1:1) medium for 5 h followed by heat shock for 1 h along with DRB in the media. Differentiation of SH-SY5Y cells was induced using 10 μM retinoic acid (Sigma) for 5 days in DMEM-F12 medium with 1% FBS. All cell lines were authenticated by short tandem repeat analysis in the UTMB Molecular Genomics Core. We routinely test mycoplasma contaminations in all our cell lines using the GeM Mycoplasma Detection Kit (Sigma) and cells were found to be free from mycoplasma contamination.

Co-immunoprecipitation (Co-IP)

Approximately 100 mg of cerebellar tissue from freshly sacrificed WT mice (6 months old; tissue from three mice were pooled together for a single experiment; n=3) were sliced into small pieces, collected in a pre-chilled, sterile homogenizer (Thomas, PHILA USA C55506) and hand-homogenized with 4 volumes of ice-cold homogenization buffer [0.25 M sucrose, 15 mM Tris-HCl, pH 7.9, 60 mM KCl, 15 mM NaCl, 5 mM EDTA, 1 mM EGTA, 0.15 mM spermine, 0.5 mM spermidine, 1 mM dithiothreitol (DTT), 0.1 mM phenylmethylsulfonyl fluoride (PMSF), and protease inhibitors (EDTA-free; Roche)] with 20 strokes to disrupt tissues (4). Homogenization was continued until a single-cell slurry was obtained, incubated on ice for 15 min, and centrifuged at 1,000×*g* to obtain the cell pellet. Nuclear extracts (NEs) were prepared as described (1, 5). Briefly, cells were lysed in Buffer A [10 mM Tris-HCl (pH 7.9), 0.34 M sucrose, 3 mM $CaCl_2$, 2 mM magnesium acetate, 0.1 mM EDTA, 1 mM DTT, 0.5% Nonidet P-40 (NP-40) and 1X protease inhibitor cocktail (Roche)] and centrifuged at 3,500×*g* for 15 min. Nuclear pellets were washed with Buffer A without NP-40 and then lysed in Buffer B [20 mM HEPES (pH 7.9), 3 mM EDTA, 10% glycerol, 150 mM potassium acetate, 1.5 mM $MgCl_2$, 1 mM DTT, 0.1% NP-40, 1 mM sodium orthovanadate (vanadate) and 1X protease inhibitors] by homogenization. Supernatants were collected after centrifugation at 15,000×*g* for 30 min and DNA/RNA in the suspension was digested with 0.15 U/µl benzonase (Novagen) at 37°C for 1 h. The samples were centrifuged at 20,000×*g* for 30 min, and the supernatants collected as NEs. Co-IPs were performed using anti-ATXN3 (MAB5360, Millipore), Lig IV (Sc-271299, Santa Cruz Biotechnology), PNKP (BB-AB0105, BioBharati Life Science) and RNAP II (920202, pSer2, H5 Ab, Biolegend) Abs with Protein A/G PLUS agarose beads (Sc 2003, Santa Cruz Biotechnology) overnight, followed by four washes with Wash buffer [20 mM HEPES (pH 7.9), 150 mM KCl, 0.5 mM EDTA, 10%

glycerol, 0.25% Triton-X-100 and 1X protease inhibitors] and eluted with Laemmli Sample Buffer (161-0737, Bio Rad; final concentration 1X). The immunoprecipitates were tested for the interacting proteins using appropriate Abs [ATXN3 (Proteintech 13505-1-AP), PNKP, 53BP1 (Sc 22760, Santa Cruz Biotechnology), DNA-PKcs (GTX 6D1 C11 F10, GeneTex), Ku70 (GTX101820, GeneTex), Lig IV (GTX108820, GeneTex), XRCC4 (GTX109632, GeneTex), Polymerase Mu (GTX116332, GeneTex), RNAP II, TFIIH (Sc 293, Santa Cruz Biotechnology), CSB (Sc 25370, Santa Cruz Biotechnology), XRCC1 (GTX111712, GeneTex), APE1 (in-house Ab) (6) and RAD51 (GTX100469, GeneTex)]. Each antibody was profiled by the respective vendor as appropriate for IP and Western blot.

For co-IP from SH-SY5Y cells, a similar protocol was followed except for the initial tissue homogenization step.

Size-exclusion chromatography

Fractionation of nuclear extract from HeLa-S3 cells was performed as described previously (7). Essentially, HeLa cell extracts were prepared by a modified Dignam method (5) and the protein complexes were separated by a Sephacryl- S300 high resolution gel filtration chromatography column (GE Healthcare, Cat # S-300HR) using AKTA pure (GE Healthcare) equipment. Fractions (1.4 ml) were collected and stored at -80ºC and subsequently analyzed. The column (1.6/60 cm, 125-ml capacity) was calibrated using molecular weight standard (blue dextran, thyroglobulin, ferretin, aldolase and conalbumin; GE Healthcare) according to the manufacturer's instructions.

Proximity Ligation Assay (PLA)

Neural stem cells (NSCs) were cultured onto 8-well chamber slides (Millicell EZ slides, Millipore), fixed in 4% paraformaldehyde in phosphate-buffered saline (PBS) for 20 min and permeabilized with 0.2% Tween 20 in PBS (PBST) for 20 min at room temperature. *In situ* protein-protein association in close proximity (~16 nm) was analyzed using a PLA (Duolink) kit as described earlier (3, 8). Anti-rabbit ATXN3 Ab (Proteintech 13505-1-AP) was used against anti-mouse Lig IV (Sc-271299, Santa Cruz Biotechnology) and RNAP II (920202, pSer2, H5 Ab, Biolegend) Abs. Images were analyzed in an AXIO Observer inverted microscope (Carl Zeiss). Nuclei were counterstained by DAPI.

DNA- and RNA-ChIP

DNA ChIP was performed from cerebellum tissue of freshly sacrificed WT mice (6 months old; n=3) as described (1, 4) with minor modifications. Briefly, 80 mg of fresh tissue was chopped into small pieces and fixed in 1% formaldehyde for 15 min with gentle agitation at room temperature to cross-link DNA to bound proteins. The sample was centrifuged at 440×*g* for 5 min at room temperature followed by the addition of 0.125 M glycine to terminate the cross-linking reaction. The sample was washed 2–3 times with ice-cold PBS (containing 1X protease inhibitor) and centrifuged each time at 440×*g* for 4 min at 4°C. The pellet was resuspended in 1 ml of ice-cold lysis buffer (10 mM EDTA, 1% (w/v) SDS, 50 mM Tris-HCl, pH 7.5 with protease inhibitor) for 15 min on ice and homogenized slowly (approximately 20 strokes on ice) with a hand homogenizer to make a single-cell suspension. After homogenization, the sample was transferred to pre-cooled 1.5 ml microcentrifuge tubes and centrifuged at 2260×*g* for 5 min. The pellet was further resuspended in ice-cold lysis buffer (500 µl) and subjected to sonication to generate average 400 bp of DNA fragments [sonicated 3 times for 1.5 min for each pulse (3-4 consecutive pulses given)].

The sample was centrifuged at 20,780×*g* for 30 min at 4°C, and the supernatant (the sheared chromatin with bound protein) was collected. The lysate was diluted with 15 mM Tris-HCl, pH 8.0, 1.0 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.01% SDS and protease inhibitors and immunoprecipitated for 4 h at 4°C with 10 µg (per mg of lysate) of isotype control IgG (Santa Cruz Biotechnology, Sc-2027) or the anti-ATXN3 Ab (MAB5360, Millipore). Immunocomplexes (ICs) were captured by Magna ChIP^{TM} Protein-A magnetic beads (Millipore) overnight and washed sequentially in buffer I (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Triton-X-100, and 0.1% SDS); buffer II (same as buffer I except containing 500 mM NaCl); buffer III (1% NP-40, 1% sodium deoxycholate, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA), and finally with 1X Tris-EDTA (pH 8.0) buffer. The ICs were extracted from the beads with elution buffer (1.0% SDS, 100 mM $NaHCO_3$), de-crosslinked for 4 h at 65°C and DNA isolated by phenol-chloroform extraction and ethanol precipitation using GlycoBlue (Life Technologies) as carrier. A similar protocol was followed for heat-shock induced ChIP studies from SH-SY5Y cells, except the initial tissue homogenization step and sonication was performed for three times for 30 seconds each.

For RNA-ChIP from SH-SY5Y cells, a similar protocol was followed as described earlier (1) with minor changes. Briefly, the crosslinking time was reduced to 10 min and nuclear extracts were prepared before sonication. The cells were incubated in buffer A (5 mM HEPES, 85 mM KCl, 0.5% NP-40 and 1X protease inhibitor) for 10 min at 4°C, then washed once with buffer B (buffer A minus NP-40) at 2,500×*g* for 5 min and the nuclear pellet was resuspended in sonication buffer [1% (w/v) SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8]. After sonication, the lysate was diluted 10X in IP buffer (15 mM Tris-HCl, pH 8.0, 1.0 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.01% SDS and protease inhibitors) and immunoreaction was carried out for 4 hrs at 4°C followed by capture of IC in protein A/G PLUS agarose beads overnight and stringency wash as described

for DNA ChIP. 50 U/ml of RNase inhibitor (Roche) was added to buffers A and B, sonication and IP buffers, and 40 U/ml to each wash buffer. De-crosslinking time was reduced to 2 h. RNA isolation was carried out in acidic phenol-chloroform followed by ethanol precipitation with GlycoBlue as a carrier. Genomic DNA was removed and reverse transcription was performed using a PrimeScript^{TM}RT Kit with gDNA Eraser (TaKaRa). ChIP samples were analyzed by qPCR/PCR using specific primers (*SI appendix,* Table S2**)**. qPCR data are represented as fold enrichment (% Input after normalization to IgG) or as fold increase in normalized % Input, in the case of heat shock experiments. For RNA-ChIP, no amplification was detected with IgG, so the % Input for IgG was taken as zero, and the data were represented as % Input for the rest of the samples.

RNase H treatment

Cells were washed with PBS, permeabilized in 2% PBST (v/v) for 10 min at room temperature, washed again with PBS, and then incubated with 150 U/ml RNase H (TaKaRa) in 1X buffer (40 mM Tris-HCl, pH 7.7, 4 mM $MgCl_2$, 1 mM DTT, 4% Glycerol) for 20 min at 37°C. The cells were then harvested, washed twice with PBS, then crosslinked with 1% formaldehyde in PBS for 10 min, and lysates were prepared for RNA-ChIP following the protocol described above.

Gene knock down by siRNA transfection

ATXN3 depletion was carried out in SH-SY5Y, HEK293, neural stem cells and BEAS-2B cells using siRNAs (80 nM; transfected twice on consecutive days) from Santa Cruz Biotechnology (Sc 40358). The cells were treated with control or specific siRNA and lipofectamine 2000 (Invitrogen) mixture for 6 hrs in optiMEM (Gibco; reduced serum media) followed by addition of 10% FBS

containing DMEM/F12 media for each round of transfection. PNKP and Lig IV depletions were carried out using Sigma siRNAs (SASI_Hs01_00067475 and SASI_Hs02_00318748, respectively). The control siRNA was purchased from Sigma (Mission universal control, SIC001). Nuclear or whole cell extracts (using RIPA buffer, Sigma) were prepared from the harvested cells (72-96 h post-transfection) to examine the depletion of individual proteins by immunoblot analysis using Abs mentioned earlier. HDAC2 (Histone deacetylase 2; GTX109642, GeneTex) or Lamin B [a gift from Dr. ML Hegde] (for nuclear extracts) and GAPDH (GTX100118, GeneTex, for whole cell extracts) were used as loading controls.

Long amplicon quantitative PCR (LA-qPCR)

The cells were mock- or Bleo-treated 72 h post ATXN3 depletion and either harvested immediately after Bleo treatment or kept for recovery (12-15 h) after the Bleo treatment and then harvested. Genomic DNA was extracted using the Genomic tip 20/G kit (Qiagen) per the manufacturer's protocol, to ensure minimal DNA oxidation during the isolation steps. The DNA was quantitated by Pico Green (Molecular Probes) in a black-bottomed 96-well plate and gene-specific LA qPCR assays were performed as described earlier (1, 4, 8, 9, 10) using Long Amp Taq DNA Polymerase (New England BioLabs). Three transcribed (Enolase, NeuroD1, TUBB, ~6 kb) and three non-transcribed (MyH2, MyH4, MyH7, ~6 kb) genes were amplified from SH-SY5Y cells and SCA3 patient and age-matched control post-mortem cerebellums (MyH6 was used as a non-transcribed gene instead of MyH7) using appropriate oligos. Similarly, two transcribed (HPRT and POLB, 10.9 and 12.1 kb, respectively) and two non-transcribed genes (NANOG, 8.6 kb and OCT3/4, 10.1 kb) were amplified from the DNA of ATXN3-depleted HEK293 cells using specific oligos. To estimate DNA damage in the heat shock gene, an 8.9 kb region of HSP70 was

amplified from genomic DNA isolated from mock- or heat shock-induced (Bleo-treated) SH-SY5Y cells.

In another experiment, brain stems, cerebellums and forebrains from freshly euthanized CMVMJD135 mice (11) (mean age of 25 weeks, n≥3), expressing human ATXN3 with 135 glutamine residues, were used for DNA damage analysis. These mice display a progressive motor phenotype starting at 6 weeks, with extensive phenotypic overlap with the human disease (11). Three transcribed (Enolase, NeuroD1, TUBB, ~6 kb) and three non-transcribed genes (MyH4, MyH6, MyoD, ~6 kb) were amplified using appropriate oligos.

Finally, genomic DNA isolated from Drosophila (20 from each genotype) adult head+thorax with motor neuron specific and adult head with eye specific expression of human ATXN3.trQ78 and ATXN3.trQ78;GS-PNKP co-expression were used for DNA damage analysis. Two genes (CrebB and Neurexin, ~8 kb) were amplified using appropriate oligos.

The LA-qPCR reaction was set for all genes from the same stock of diluted genomic DNA sample, to avoid variations in PCR amplification during sample preparation. Preliminary optimization of the assays was performed to ensure the linearity of PCR amplification with respect to the number of cycles and DNA concentration (10-15 ng). The final PCR reaction conditions were optimized at 94°C for 30 s; (94°C for 30 s, 55-60°C for 30 s depending on the oligo annealing temperature, 65°C for 10 min) for 25 cycles; 65°C for 10 min. Since amplification of a small region is independent of DNA damage, a small DNA fragment (~200-400 bp) from the corresponding gene(s) was also amplified for normalization of amplification of the large fragment. The amplified products were then visualized on gels and quantitated with ImageJ software (NIH). The extent of damage was calculated in terms of relative band intensity with a control siRNA/mock-treated

sample or WT mice/control human subject or OK6/GMR control Drosophila considered as 100. All oligos used in this study are listed in *SI appendix*, Table S2.

DSB repair plasmid assay in *lacZ*-expressing stable HEK293 cell line

We described earlier in detail the generation of a stable WT *lacZ*-expressing HEK293 cell line, and the construction of a DSB-containing plasmid for *in cell* repair assay (1). The assay employs a plasmid (pTV123) expressing functional *lacZ* while carrying a silent mutation within the *lacZ* gene, to distinguish its transcript from the endogenous WT *lacZ* transcript in the HEK293 stable cell line. The WT *lacZ* transcript serves as a template for transferring the missing sequence into the gapped plasmid pTV123. Briefly, the plasmid pTV123, isolated from *E.coli dam*$^{-}$*dcm*$^{-}$ strain (NEB) was digested separately with the two single-cut enzymes *Bsa*BI and *Bcl*I (NEB) bracketing the silent mutation, to produce 3'-OH ends containing DSBs in the plasmid. In another case, two complementary HPLC-purified uracil (U)-containing oligos (*SI appendix*, Table S2) with the same silent mutation in the *lacZ* sequence were annealed and ligated to the *Bsa*BI/*Bcl*I-digested vector. To create DSB-containing 3'-P ends, the resulting plasmids were digested with Udg and Fpg (both from NEB) and the linear form of the plasmid was isolated from the agarose gel.

To assess the role of ATXN3 in RNA-mediated DSB repair, HEK293 stable *lacZ*-expressing cells were initially transfected with ATXN3-specific or control siRNA and kept for 48 h. The cells (50-60% confluent) were then further transfected with 250 ng of the DSB-containing plasmid (3'-OH or 3'-P ends) using Lipofectamine 2000, and kept for 16 h following transfection to allow repair. Plasmids were recovered and DH5α *recA*$^{-}$ *lacZ*$^{-}$ cells (NEB) were transformed with the recovered DNA as previously described (1). The blue colonies, selected on Ampicillin (50 µg/ml)/X-gal (40 µg/ml) agar plates, were counted and the presence of WT *lacZ* was confirmed by sequencing

(UTMB Molecular Genomics Core). For control experiments, PNKP or Lig IV were depleted individually and DSB-containing plasmid with 3'-OH end was used for transfection.

To determine the expression level of *lacZ* transcript following ATXN3 depletion, RNA was isolated using the RNeasy Mini Kit (Qiagen), genomic DNA was removed and reverse transcription was performed using a PrimeScript^{TM}RT Kit with gDNA Eraser (TaKaRa). qRT-PCR was performed using SYBR green mix (TaKaRa) with *lacZ* gene-specific oligos (*SI appendix*, Table S2) with the GAPDH gene used as housekeeping control.

Human tissue samples

Human post-mortem cerebellum tissue from SCA3 patients and age-matched controls (*SI appendix*, Table S1) were obtained from the bio repository of the Michigan Brain Bank, USA.

3'-phosphatase activity of PNKP

The 3'-phosphatase activity of PNKP in the nuclear extract of WT vs. SCA3 mice (0.5 µg) brainstem/forebrain and post-mortem patient cerebellums and age-matched control subjects (2.5 µg) or with purified recombinant PNKP (25 fmol) was conducted as we described previously (6, 10, 12). Five pmol of the radiolabeled substrate was incubated at 37°C for 15 min in buffer A (25 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 10% glycerol and 0.1 µg/µl acetylated BSA). Nuclear extracts were prepared following the protocol used for Co-IP studies. The radioactive bands were visualized in PhosphorImager (GE Healthcare) and quantitated using ImageQuant software. The data were represented as % product (phosphate) released from the radiolabeled substrate with a value arbitrarily set at 100%.

The Drosophila homolog of human *PNKP*, CG9601 was cloned in pET24d+ vector (Addgene), expressed in *E. coli* BL21(DE3) Codon Plus RIL and purified using affinity chromatography with Cobalt resin (Fischer Scientific). The purified protein (10-20 ng) was assayed for PNKP's phosphatase activity as described above.

Immunoblotting

The proteins in the nuclear extracts/whole cell extracts were separated onto a Bio-Rad 4-20% gradient Bis-Tris Gel, then electro-transferred on a nitrocellulose (0.45 μm pore size; GE Healthcare) membrane using 1X Bio-Rad transfer buffer. The membranes were blocked with 5% w/v skimmed milk in TBST buffer (1X Tris-Buffered Saline, 0.1% Tween 20), then immunoblotted with appropriate antibodies [PNKP, RNAP II, ATXN3, γ-H2AX (ser phos 139 residue; GTX628789, GeneTex), p53BP1 (S1778, 2675S Cell Signaling Technology), 53BP1, HDAC2, NEIL1 (in house) (6) and NEIL2 (in-house) (4)]. The membranes were extensively washed with 1% TBST followed by incubation with anti-isotype secondary antibody (GE Healthcare) conjugated with horseradish peroxidase in 5% skimmed milk at room temperature. Subsequently, the membranes were further washed three times (10 min each) in 1% TBST, developed using ECL™ Western Blotting Detection Reagents (RPN2209, GE Healthcare) and imaged immediately.

For assessing ATXN3.trQ78 and RNAP II levels, 15 fly (2-days-old) heads per genotype were collected, lysed in Laemmli buffer by sonication, boiled at 95°C for 5 min, and centrifuged at full speed for 10 min to collect the supernatant. Samples were separated on 4-12% SDS-PAGE gels and transferred to PVDF membranes (Millipore, Fisher Scientific). After blocking in 5% (w/v) non-fat dried milk in PBST for 1 h, the membrane was incubated with primary antibody [ATXN3,

RNAP II or mouse anti-actin (mAbGEa, Cat # 74340)] overnight at 4°C. After incubation with secondary antibody, ECL (Amersham) signal was detected in a ChemiDocTM Touch (Bio-Rad). For the SCA3 motor neuron model, larval brains (20 each) were used for Western analysis to assess the RNAP II level.

Immunofluorescence (IF) microscopy

Neural stem cells (NSC) were cultured in 8 well chamber slides (Millipore), fixed in 4% paraformaldehyde in PBS for 20 min and permeabilized with 0.2% Tween 20 in PBS for 20 min at room temperature. Blocking was performed with 3% BSA solution in DPBS for 30 min. Incubation was done with γ-H2AX, p53BP1 and RAD51 [14B4,GTX70230, GeneTex) Abs overnight at 4°C in 1% BSA solution and fluorescent secondary antibodies for 1 h at 37°C in 1% BSA solution. Slides were washed three times and counterstained with DAPI. Images were captured by an AXIO Observer inverted microscope (Carl Zeiss).

Ubiquitination of RNAP II large subunit in ATXN3 depleted cells/SCA3 mice tissue

The control and ATXN3-siRNA transfected cells were treated for 6 h with cell permeable proteasome inhibitor MG-132 (Sigma, final concentration 10 µM), 72-96 h post siRNA transfection and the nuclear extracts (NE) were prepared from the harvested cells. To maintain denaturing condition, we used the following nuclear lysis buffer composition: 50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 5 mM $MgCl_2$, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS, 20 mM N-Ethylmaleimide (NEM), 1X complete EDTA free protease inhibitor (Roche), supplemented with 0.15 U/µl benzonase (Novagen). The same buffer was used for NE preparation from mouse cerebellum with the supplementation of 10 µM MG132 in the buffer itself. NEs were

IP'ed with anti-RNAP II (H5) Ab and probed with the anti-ubiquitin antibody (3933S, Cell Signalling Technology) to assess the ubiquitination level of RNAP II.

RNA synthesis assay

ATXN3 or control siRNA transfected cells were grown in a 4 well chamber slide overnight and 5-ethynyl-uridine (EU) was added directly to the culture medium followed by incubation for 2 h to determine the level of nascent RNA. Incorporation of EU was visualized by *Click IT* conjugation of Alexa Fluor 488 according to the manufacturer's protocol (Invitrogen, Cat# 10329). Images were obtained using a Zeiss LSM 710 confocal microscope. Fluorescence-signal intensities were quantified using the ImageJ software (NIH). Fifty cells were analyzed randomly and background was subtracted for each experiment. UV (20 J/m^2)-exposed cells were used as positive control.

Total repair assays

DNA single-strand break repair assays (containing either 3'-P or 3'-OH ends) were carried out essentially as we described earlier (6, 8, 10) with minor modifications. Briefly, nuclear extracts (2.5 μg) of SCA3 patients vs. age-matched controls were incubated with 10 pmol of a 51-mer DNA-oligo (Integrated DNA Technologies) annealed to two other 25-mer DNA oligos containing either 3'-P and 5'-P (substrate for PNKP-mediated end processing followed by DNA polymerase mediated gap filling and ligase activities) or 3'-OH and 5'-P (no PNKP activity is required but polymerase and ligase activities are required for total repair) with a single nucleotide gap in the middle (*SI appendix*, Table S2**)**. The reaction mixture (20 μl) contained 1 mM ATP, 50 μM unlabeled dNTP, and 0.5 pmol [α-^{32}P] dCTPs (the concentration of the corresponding cold dNTP was lowered to 5 μM) in BER buffer (6, 10) and the reaction mixture was incubated for 45 min at 30°C. The reaction products were run in a denaturing 20% urea-PAGE and the radioactive bands

were detected in a PhosphorImager (GE Healthcare). The data were represented as % of the ligated product (% repair) with the sum of both un-ligated and ligated products arbitrarily considered as 100.

DNA glycosylase/AP lyase assay for the NEILs

DNA glycosylase and AP lyase activity of NEIL1 and 2 in the nuclear extract of SCA3 vs. WT mice (1.0 µg) and SCA3 patient vs. age-matched controls (2.5 µg) were measured using a 5'-^{32}P-labeled 51-mer oligo with 5-OHU at position 26 from the 5′-end (Midland Certified Reagent; *SI appendix*, Table S2) in an 11 nucleotide long bubble as described earlier (13). The data were then represented as % of the 25-mer cleaved product (% product) with the sum of both labeled 51-mer and 25-mer cleaved product arbitrarily considered as 100.

Analysis of gene expression by real time qRT-PCR

For analysis of gene expression in the SH-SY5Y cell line (ATXN3 depleted), SCA3 mouse and post-mortem human patient subjects, RNA was extracted from cells/tissues (brain stem or cerebellum) and cDNA was prepared as described earlier. qRT PCR was performed using gene specific primers as listed in *SI appendix*, Table S2.

Fly stocks and rearing conditions

All fly lines were maintained at 25°C on standard fly food under a 12:12-h light–dark cycle. The following stocks were obtained from the Bloomington Drosophila stock center (BDSC, Bloomington, Indiana, USA): w^{1118} (#3605), UAS-ATXN3.trQ78 (#8150), OK6-Gal4 (#64199) and GMR-Gal4 (#1104). The GS-CG9601 line (stock # 202883) was obtained from the Drosophila Genetic Resource Center (Kyoto, Japan).

Adult fly climbing assay (Negative-geotaxis)

For each group, a maximum of 20 adult flies (usually 5-15) per vial were analyzed. Flies were age-matched (hatched within a 2-days window) and pre-mated. We used male flies because female flies exhibit more variability in the behavior. The night before the assay, flies were transferred to a new tube to allow grooming of food debris. The assay is run at approximately the same time in the day, with the same setting of ambient light. The climbing vial was divided into six compartments, with the lowest compartment numbered with 1 (slowest climbers) and the highest with number 6 (fastest climbers). The vial was placed against a white surface and a white background to facilitate photography. Flies were transferred from their housing vial to the climbing vial, covered with a plastic plate on top. Flies were gently tapped down to the bottom of vials, and were allowed 10s to climb up, after which a picture was taken. The numbers of flies in each compartment are counted for each time point, and recorded on the worksheet. Each cohort of flies was tested 5 trials consecutively, allowing about 1 min rest in between each trial. The average score of each cohort was determined by dividing the total score by the total number of flies.

Real-time quantitative PCR (RT-qPCR)

About 50 flies were collected to isolate total RNA using the RNeasy Mini Kit (Qiagen). iScript cDNA synthesis kit (Bio-Rad #1708891) was used to transcribe RNA into cDNA. Relative quantification of the gene expression level was determined in CFX ConnectTM (Bio-Rad) by using iQTM SYBR® Green Supermix (Bio-Rad Laboratories, Inc.). Gene expression levels were normalized to a housekeeping gene, ribosomal protein 49 (RP49). Oligo sequences are provided in *SI appendix*, Table S2.

Immunohistochemistry of Drosophila

Wandering third instar larvae were dissected in $Ca2^{+}$-free ice-cold saline and fixed in 4% paraformaldehyde for 20 min. After washing with PBST (0.05% Triton-X), samples were blocked in 10% normal goat serum for 30 min, then incubated in primary antibodies overnight at 4°C, and in secondary antibodies for 1–2 h at room temperature, then mounted in VectaShield (Vector Laboratories) for microscopic analysis. We used the following antibodies: Anti-histone H2A.v pSer137 antibody (Acris, AP09307PU-N, diluted 1:200) and Alexa Fluor 488- or 568-conjugated secondary antibody (diluted 1:500). Confocal images were obtained on a microscope (Zeiss LSM 880) with fast AiryScan, 63x oil objective lens and the accompanying software. Images were processed in the Zen image browser (Zeiss). Brightness and contrast were adjusted using FIJI (ImageJ). All images were captured at room temperature (21–23°C).

Statistical analysis

Unpaired Student's *t*-test (http://www.ruf.rice.edu/~bioslabs/tools/stats/ttest.html) was used for analysis of statistical significance between two sets of data, unless mentioned otherwise. Significance was evaluated at level $P<0.05$ (*), $P<0.01$ (**) and $P<0.005$ (***), as appropriate. Climbing score of flies were assessed by Tukey's multiple comparison test with mixed effect analysis (*$p<0.05$). Similarly, Climbing score of flies at 3 weeks of age was compared by Sidak's multiple comparison test with one-way ANOVA (*$p<0.05$).

Tables

Supplementary Table S1: Details of the human post-mortem brain tissues (cerebellum)

Brain Bank ID	Age of Death	Sex	PMI (hours)	Neuropathology
1035	59	Female	4	SCA Type 3
1547	84	Female	40	SCA Type 3
1602	48	Female	22	SCA Type 3
1832	49	Male	48	SCA Type 3
1073	47	Male	23	Control
409	48	Female	5	Control
729	59	Male	12	Control
1432	83	Female	21	Control

Supplementary Table S2: Primers used in the study

Primers	Gene	Nucleotide sequence 5′ - 3′	Purpose
TV001	n/a	GCTTAGCTTGGAATCGTATCAT GTA(P)	Total DNA repair (3'P-25 mer oligo)
TV002	n/a	(P)ACTCGTGTGCCGTGTAGACC GTGCC	Total DNA repair (5'P-25 mer oligo)
TV003	n/a	GCTTAGCTTGGAATCGTATCAT GTA	Total DNA repair (3'OH-25 mer oligo)
TV031	n/a	GGCACGGTCTACACGGCACAC GAGTGTACATGATACGATTCC AAGCTAAGC	Total DNA repair (DNA 51 mer oligo)
TV140-F	POLB	AGTGGGCTGGATGTAACCTG	SA-PCR
TV141-R	POLB	CCAGTAGATGTGCTGCCAGA	SA-PCR
TV142-F	HPRT	TGGGATTACACGTGTGAACCAAC C	LA-qPCR
TV143-R	HPRT	GCTCTACCCTCTCCTCTACCGT CC	LA-qPCR
TV144-F	HPRT	TGCTCGAGATGTGATGAAGG	SA-PCR
TV145-R	HPRT	CTGCATTGTTTTGCCAGTGT	SA-PCR
TV161-F	POLB	CATGTCACCACTGGACTCTGC AC	LA-qPCR
TV162-R	POLB	CCTGGAGTAGGAACAAAAATT GCT	LA-qPCR
TV199-F	*E.coli lacZ*	GGCGACTTCCAGTTCAACATC A	Sequencing
TV200-F	*E.coli lacZ*	CATTATCCGAACCATCCGCTGT G	PCR, qPCR and Sequencing
TV201-R	*E.coli lacZ*	CTGGTCTTCATCCACGCGCGCG	PCR, qPCR and Sequencing
TV205-R	*E.coli lacZ*	CGCTATGACGGAACAGGTATT	RT-PCR

TV206-R	*E.coli lacZ*	CCGCATCAGCAAGTGTATCT	PCR from gDNA and Sequencing
TV209-F	*E.coli lacZ*	GTGCCGGCTAGCTGGCTGGAG TGCGATCTTCCT	Sequencing; RT-PCR
TV210-R	*E.coli lacZ*	TCCTGGGGATCCAGACATGAT AAGATACATT	Construction of pTV108 and PCR from gDNA
TV213-F	*E.coli lacZ*	CGTAGCTAGCCAGGCTGGGAC ACT	Construction of pTV108 and PCR from gDNA
TV214-F	NANOG	CTCCGGAATGGTAGTCTGAGA AGAA	LA-qPCR
TV215-R	NANOG	ATTTAGGGCAGGCACAAGATG G	LA-qPCR
TV218-F	HSP70	GTCAACATGGTGAAATCCCGT CTCTACTA	LA-qPCR
TV219-R	HSP70	TAGGAAATGCAAAGTCTTGAA GCTCCAAA	LA-qPCR
TV232-F	OCT3/4	TCTGTGGCCTCACCCTATGA	LA-qPCR
TV233-R	OCT3/4	CAGACCTGTGGCAGGTATTGA A	LA-qPCR
TV263-F	*E.coli lacZ*	(P)UAATCACCC**c**AGUGT	DSBR plasmid assay
TV264-R	*E.coli lacZ*	(P)GATCACACUGGGGUGATTA	DSBR plasmid assay
TV286-R	*E.coli lacZ*	GCCACATATCCTGATCTTCCAG AT	Sequencing; RT-PCR
TV287-F	*E.coli lacZ*	CTCTGGATGTCGCTCCACCAG GTAAACAG	Sequencing
TV07 RPmix	NANOG	RealTimePrimers.com	SA-PCR
TV08 RPmix	HSP70	RealTimePrimers.com	SA-PCR
TV10 RPmix	OCT3/4	RealTimePrimers.com	SA-PCR
TV19 RPmix	GAPDH (Human)	RealTimePrimers.com	Internal control for gene expression
TV21 RPmix	NeuroD1 (Human)	RealTimePrimers.com	SA-PCR and qRT-PCR
TUBB-In2-For	Tubβ	TCCCTGTCTCCCACTTATCTG	RNA-ChIP PCR/qPCR
TUBB-In2-Rev	Tubβ	TCTGGCAGAAGGGAAGGTTT	RNA-ChIP PCR/qPCR
HPRT1-In2-For	HPRT	TGCCAGTATGGGTGGGAGAA	RNA-ChIP PCR/qPCR
HPRT1-In2-Rev	HPRT	AGGGTAAAAACCCAGGCATGA	RNA-ChIP PCR/qPCR
PolB-int1-For	POLB	CCTGCCCTTAGCCCTCTTTT	RNA/DNA-ChIP PCR/qPCR
PolB-int1-Rev	POLB	AAGGAGGGAGCACAATCAGC	RNA/DNA-ChIP PCR/qPCR
HSP 70 FP	HSP70	CTGTGCGGCTGCAGGCACCGG C	DNA ChIP qPCR
HSP 70 RP	HSP70	TGGTGCGGTTGCCCTGGTCGTT	DNA ChIP qPCR
HSP89 FP	HSP89	CCTCTGTAGACGTCCTGCAAG GT	DNA ChIP qPCR
HSP89 RP	HSP89	ATCCGATTCTGGGTTAATAAGT G	DNA ChIP qPCR
18S Primer mix	18S (Human)	RealTimePrimers.com	Endogenous control in qRT-PCR.
H-TUBB3-LA-F	TUBB (human)	TGCTTCTCATGCTTGCTACCAC	LA-qPCR
H-TUBB3-LA-R	TUBB (human)	TCTGTCCCTGTAGGAGGATGT	LA-qPCR

H-NeuroD1-LA-F	NeuroD1 (human)	CCGCGCTTAGCATCACTAAC	LA-qPCR
H-NeuroD1-LA-R	NeuroD1 (human)	TGGCACTGGTTCTGTGGTATT	LA-qPCR
H-ENO2-LA-F	Enolase (human)	ACGTGTGCTGCAAGCAATTT	LA-qPCR
H-ENO2-LA-R	Enolase (human)	CCTGAAACTCCCCTGACACC	LA-qPCR
H-MyH2-LA-F	MyH2 (human)	AAAGCCTGCCAAGCCCTAAA	LA-qPCR
H-MyH2-LA-R	MyH2 (human)	TGGTCAGCATGGCAAGTGAA	LA-qPCR
H-MyH4-LA-F	MyH4 (human)	CAGGAGTGGTCCCTAAAGGC	LA-qPCR
H-MyH4-LA-R	MyH4 (human)	GTAAAAACACGGTCCCTGCC	LA-qPCR
H-MyH6-LA-F	MyH6 (human)	ATTGACTCGGTGCCCTTTCT	LA-qPCR
H-MyH6-LA-R	MyH6 (human)	TGCCACCCAACATGGGTATT	LA-qPCR
H-MyH7-LA-F	MyH7 (human)	GCTGCTCCTCTGGTTAAGGG	LA-qPCR
H-MyH7-LA-R	MyH7 (human)	GGGAGCCTCAGTCCCTACTT	LA-qPCR
H-TUBB Primer mix	TUBB (human)	RealTimePrimers.com	SA-PCR/qRT-PCR
H-Enolase Primer mix	Enolase (human)	RealTimePrimers.com	SA-PCR/qRT-PCR
H-MyH2 Primer mix	MyH2 (human)	RealTimePrimers.com	SA-PCR
H-MyH4 Primer mix	MyH4 (human)	RealTimePrimers.com	SA-PCR
H-MyH6 Primer mix	MyH6 (human)	RealTimePrimers.com	SA-PCR
H-MyH7 Primer mix	MyH7 (human)	RealTimePrimers.com	SA-PCR
m-*TUBB*-LA-F	*TUBB* (mouse)	GGTACAGGGGATGTGGTTGG	LA-qPCR
m-*TUBB*-LA-R	*TUBB* (mouse)	GAGTCTCCTGCCTGTCCCTA	LA-qPCR
m-*ENO2*-LA-F	*Enolase* (mouse)	CGGAGAGAGTTACCGAGTAC	LA-qPCR
m-*ENO2*-LA-R	*Enolase* (mouse)	TCCCTCCTAAGGAACACAGCA	LA-qPCR
m-*NeuroD1*-LA-F	*NeuroD1* (mouse)	CTCGCAGGTGCAATATGAATC	LA-qPCR
m-*NeuroD1*-LA-R	*NeuroD1* (mouse)	GCAACTGCATGGGAGTTTTCT	LA-qPCR
m-*MyH4*-LA-F	*MyH4* (mouse)	GACGTGGAACTGTTAGGCCA	LA-qPCR
m-*MyH4*-LA-R	*MyH4* (mouse)	AAGCCAGAGTCTTCAACCCG	LA-qPCR
m-*MyH6*-LA-F	*MyH6* (mouse)	GACAAGGGGCATTGTAGCCT	LA-qPCR

m-*MyH6*-LA-R	*MyH6* (mouse)	TCTGCCTACCTTATGGGGCT	LA-qPCR
m-*MyoD*-LA-F	*MyoD* (mouse)	ATAGACTTGACAGGCCCCGA	LA-qPCR
m-*MyoD*-LA-R	*MyoD* (mouse)	GGACCGTTTCACCTGCATTG	LA-qPCR
m-*TUBB* Primer mix	*TUBB* (mouse)	RealTimePrimers.com	SA-PCR/qRT-PCR
m-*Enolase* Primer mix	*Enolase* (mouse)	RealTimePrimers.com	SA-PCR/qRT-PCR
m-*NeuroD1* Primer mix	*NeuroD1* (mouse)	RealTimePrimers.com	SA-PCR/qRT-PCR
m-*MyH4* Primer mix	*MyH4* (mouse)	RealTimePrimers.com	SA-PCR/qRT-PCR
m-*MyH6* Primer mix	*MyH6* (mouse)	RealTimePrimers.com	SA-PCR/qRT-PCR
m-*MyoD* Primer mix	*MyoD* (mouse)	RealTimePrimers.com	SA-PCR/qRT-PCR
m-*gapdh* Primer mix	*gapdh* (mouse)	RealTimePrimers.com	SA-PCR/qRT-PCR
5-OHU-51	n/a	GCT TAG CTT GGA ATC GTA TCA TGT A*X*A CTC GTG TGC CGT GTA GAC CGT GCC (*X*=5OHU)	5-OHU containing 51-mer oligo
B11 Complement	n/a	GGC ACG GTC TAC ACG GCA CAA ACA GCC CAC GGA TAC GAT TCC AAG CTA AGC	Complementary to 5-OHU-51 oligo to form 11-nt long bubble
RP49 FP	RP49 (Drosophila)	CCGCTTCAAGGGACAGTATC	RT-qPCR
RP49 RP	RP49 (Drosophila)	GACAATCTCCTTGCGCTTCT	RT-qPCR
CG9601 FP	PNKP (Drosophila)	GTTTCTTCGTCGGCGATGCAG	RT-qPCR
CG9601 RP	PNKP (Drosophila)	CCCACGTTGGCTGCAAAAAG	RT-qPCR
DmCrebB LA F	CrebB (Drosophila)	TCGAGTGTCAGGTGTTACGG	LA-qPCR
Dm CrebB LA R	CrebB (Drosophila)	TCGCCACTGTAGTGAAGAGG	LA-qPCR
Dm CrebB SA F	CrebB (Drosophila)	AGATCCTGCGGAGTCTACGA	SA-PCR
Dm CrebB SA R	CrebB (Drosophila)	GGCGAGGCTCACTAATCTGG	SA-PCR
DmNeurexin LA F	Neurexin (Drosophila)	GGTACACCGGACTGAATGGG	LA-qPCR
DmNeurexin LA R	Neurexin (Drosophila)	CTGGCTACACTTGGTGGGTC	LA-qPCR
DmNeurexin SA F	Neurexin (Drosophila)	TGCAACAACAGTGCCCTACT	SA-PCR
DmNeurexin SA R	Neurexin (Drosophila)	TAGCCTTAACGAGCGACCAC	SA-PCR

Supplementary Figure Legends

Fig. S1: ATXN3 associates with RNAP II and C-NHEJ proteins

(A) Upper Panel: Nuclear extracts (NEs) (benzonase-treated) from SH-SY5Y cells (either mock- or Bleo-treated) were immunoprecipitated (IP'ed) with anti-ATXN3 Ab (lanes 4, 5) or control IgG (lanes 6, 7) and tested for the presence of associated proteins with specific Abs as indicated to the right of each row. IP experiment was repeated at least three times from separate batches of cells, and one representative figure is shown. **Lower panel:** The quantitation of the IP'ed bands after normalization with respective ATXN3 band intensity (n=3). The normalized band intensity of the mock treated samples arbitrarily considered as unity. Error bars represent ±SD of the mean. The increase in IP band intensity following Bleomycin treatment was significant at the level of *** =P<0.005, ** =P<0.01.

(B) HeLa-S3 cell nuclear extracts were fractionated on a Sephacryl S-300 gel filtration column. Immunoblot shows that a significant fraction of ATXN3 elutes as a megadalton-size complex (1.5 MDa) (panel within red box), which contains RNAP II, PNKP and Lig IV.

Fig. S2: Role of ATXN3 in C-NHEJ and RT-PCR profiling of the mouse and human genes

(A) Upper Panel: γ-H2AX foci formation in ATXN3-depleted neural stem cells. Nuclei were counterstained with DAPI (blue). **Lower Panel**: The average number of γ-H2AX foci per cell is shown in the bar diagram, calculated from 20 randomly selected cells/ sample. Error bars represent ±SD of the mean. The increase in foci/cell in ATXN3-depleted cells was significant at the level of *** =P<0.005.

(B) Upper Panel: p53BP1 (red) and RAD51 (green) foci formation in ATXN3-depleted post-mitotic, non-cycling SH-SY5Y cells (following differentiation by retinoic acid). **Lower Panel**:

Nuclei were counterstained with DAPI (blue). The average number of foci per cell is shown in the bar diagram, calculated from 20 randomly selected cells/ sample. Error bars represent ±SD of the mean. The increase in p53BP1 foci/cell in ATXN3-depleted cells was significant at the level of *** =P<0.005.

Gene expression profile of various mouse and human genes used in the study. The PCR fragments shown here were obtained from cDNA generated by RT reaction with total RNA extracted from **(C)** Cerebellum of WT mice or **(D)** SH-SY5Y cells.

To rule out the possibility of gDNA contamination, no RT enzyme was added to "-RT" samples. Constitutively expressing mouse and human GAPDH were used as positive controls.

Fig. S3: ATXN3 depletion and evaluation of genomic strand-break levels in transcribed vs. non-transcribed genes by LA-qPCR

(A) Left Panel: Western blots showing the levels of proteins (as indicated in the right of each row) in the nuclear extract of ATXN3-depleted SH-SY5Y cells. HDAC2: used as nuclear loading control. **Right Panel:** shows the quantitation of ATXN3 depletion and concurrent decrease in RNAP II level (***=P<0.005).

(B) Western blots showing the levels of proteins (as indicated to the right of each row) in the nuclear extracts of ATXN3-depleted HEK293 cells. HDAC2: used as nuclear loading control.

(C) HEK293 cells were transfected with either control siRNA (lanes 1-3) or ATXN3 siRNA (lanes 4-6) and further mock- (-) or Bleo- (+) treated or kept for recovery (+/R) after Bleo treatment, and harvested for genomic DNA isolation.

Upper Panel: Representative agarose gel images of each long (~8-12 kb) and short (~200-400 bp) amplicon of **(C)** the transcribed (HPRT and POLB) genes and **(D)** the non-transcribed (NANOG

and OCT3/4) genes. **Lower Panel:** The bar diagram represents the normalized data as relative band intensity, with the control siRNA/mock-treated sample (lane 1) arbitrarily set as 100. n≥3; Error bars represent ± SD of the mean. The persistence of damage after recovery for each transcribed gene (**C**, lane 6) was significant (***=P<0.005) compared with the corresponding mock-treated samples (**C**, lane 4); however, the damage was almost completely repaired in non-transcribed genes [**D**, lane 6 vs. lane 4; ns=non-significant (P>0.05)].

Fig. S4: Association of pre-mRNA with ATXN3

(A) The RNA-ChIP assays were performed from mock- or Bleo- or GO-treated SH-SY5Y cells using anti-ATXN3 Ab. Mock reverse transcription reaction without RT enzyme was performed from IP'ed samples to rule out genomic DNA contamination. RT-PCR was performed for Tubß, HPRT and POLB genes using intron-specific primers.

(B) Similar control experiment for the time course analysis.

(C) Similar control experiment for the RNase H treatment**.**

Western blots showing the levels of proteins (as indicated in the right of each row) in the whole cell extract of **(D)** ATXN3-depleted HEK293-Pcmv^{+}*lacZ* stable cells. **Right Panel** for **(D):** shows the quantitation of ATXN3 depletion and concurrent decrease in RNAP II level (***=P<0.005). GAPDH was used as loading control.

(E) Relative expression profile of the *lacZ* gene in Control vs. ATXN3-depleted HEK293-Pcmv^{+}*lacZ* stable cells. GAPDH was used as the loading control (ns=non-significant, =P>0.05).

Western blots showing the levels of proteins (as indicated in the right of each row) in the whole cell extract of **(F)** PNKP-depleted **(G)** Lig IV-depleted HEK293-Pcmv^{+}*lacZ* stable cells.

Fig. S5: PNKP/NEIL activity and LA-qPCR in the WT vs. SCA3 mice

(A) Upper Panel: PNKP mediated 3'-phosphatase activity in the nuclear extract of WT (lanes 2-4) vs. SCA3 mice (lanes 5-7) forebrain. Lane 1: No protein, substrate only. Lane 10: purified PNKP (25 fmol) as positive control. **Lower Panel**: Quantitation of the products (released phosphate) is represented in the bar diagram (n=3, ns= non-significant, P>0.05). Error bars represent ±SD of the mean.

(B) Upper Panel: A 5′ ^{32}P-labeled 51-mer oligo (5-OHU.B11, 10 pmol) was used for DNA glycosylase/AP lyase activity of the NEILs (NEIL1 and 2) in the nuclear extracts (1.0 µg) of WT (lanes 2-5) vs. SCA3 (lanes 6-9) mice brain stem. Lane 1: no protein; lane 10: purified NEIL2 (50 ng). **Lower Panel:** Bar diagram represents the quantitation of the excised 25-mer oligo product formed (n=4).

(C) Upper Panel: Genomic DNA was isolated from 25-weeks-old WT and SCA3 transgenic mice forebrain. Amplification of each long amplicon (6-8 kb) and a short amplicon (~200 bp) of the transcribed (Enolase, NeuroD1 or TUBB) genes. **Lower Panel**: the normalized data are represented in the bar diagram as relative band intensity with the WT sample arbitrarily set as 100 (from three replicate gels one representative gel is shown; error bars represent ±SD of the mean), ns= non-significant (P>0.05).

Fig. S6: ATXN3 specifically blocks PNKP's activity

Total DNA repair synthesis was performed with the nuclear extract (2.5 µg) isolated from age-matched healthy normal (control) vs. SCA3 patients' cerebellum tissue using two nicked DNA duplexes:

(A) One with 3'-OH ends that do not require PNKP activity but need DNA polymerase and ligase activities for repair completion.

(B) The other with 3'-P termini that require PNKP's 3'-phosphatase activity along with DNA polymerase and DNA ligase activities for repair completion.

(Upper Panels) in **(A)** and **(B):** schematic representations of the assays. **(Middle Panels)** in **(A)** and **(B)**: The 51-mer upper DNA band represents repaired DNA duplexes. Lane M represents 51-mer (upper) and 25 mer (lower) markers. **(Lower Panels)** in **A** and **B:** Bar diagrams showing relative band intensity of the repaired DNA product with the total repair in each age-matched control arbitrarily set as 100. Data represent mean ± SD, n=3, ***=P<0.001.

(C) NEIL (NEIL1 and 2) glycosylase/AP lyase activity in the nuclear extracts (2.5 µg) of age-matched healthy normal (control) vs. SCA3 patients. Lane 1: no protein; lane 10: purified NEIL2 (50 ng). **Upper Panel:** One representative gel is shown. **Lower Panel:** Bar diagram represents the quantitation of the excised 25-mer oligo product formed (n=3).

(D) Western blots showing the levels of proteins (indicated to the right of each row) in the NE of age-matched healthy normal (control) vs. SCA3 patients' cerebellum.

Fig. S7: Effect of ATXN3-depletion on global transcription

(A) Upper Panel: Western blots showing the levels of proteins (as indicated in the right of each row) in the nuclear extracts of ATXN3-depleted neural stem cells. **Lower Panel:** Quantitation of ATXN3 depletion and concurrent decrease in RNAP II level (n=3, ***=P<0.005). HDAC2: used as nuclear loading controls.

(B) Similar experiment in BEAS-2B cells. (n=3, ***=P<0.005, **=P<0.01). Lamin B: Used as nuclear loading control.

(C) BEAS-2B cells were transfected with control or ATXN3 siRNA and the nuclear extracts were immunoprecipitated (IP'ed) with IgG (lanes 1, 2) or RNAP II Abs (elongating form, H5; lanes 3, 4) followed by Western blotting with anti-Ubiquitin Ab to assess the ubiquitination level of RNAP II. The upper Arrow indicates slower migrating poly-ubiquitinated forms of RNAP II [$(Ub)_n$-RNAP II]. The lower arrow indicates the expected position of non-ubiquitinated RNAP II. The positions of the protein markers are indicated on the left.

(D) Nuclear extracts from WT and SCA3 mice cerebella were immunoprecipitated (IP'ed) with IgG (lanes 1, 2) or RNAP II Abs (elongating form, H5; lanes 3, 4) followed by Western blotting with anti-Ubiquitin Ab to assess the ubiquitination level of RNAP II. The upper Arrow indicates slower migrating poly-ubiquitinated forms of RNAP II [$(Ub)_n$-RNAP II]. The lower arrow indicates the expected position of non-ubiquitinated RNAP II. The positions of the protein markers are indicated on the left.

(E) EU labeled RNA was detected by confocal microscopy (10X magnification) in control siRNA or ATXN3 siRNA transfected BEAS-2B cells. EU incorporation following UV exposure (20 J/m^2) was used as control. Bar diagram represents the relative fluorescence intensity (from 50 randomly chosen cells) as a measure of the *de novo* transcript level of the ATXN3-depleted cells with the intensity in control siRNA-transfected cells arbitrarily set as 100 (***=P<0.005).

Fig. S8: Analysis of expression level of critical neuronal cell-specific genes

Expression profile of genes critical for neuronal functions (TUBB, Enolase, NeuroD1) in Control vs. ATXN3-depleted **(B)** SH-SY5Y cells **(A)**, WT vs. SCA3 mouse brain stem **(C)** and cerebellum **(D)** and control vs. SCA3 post-mortem patients (age-matched) cerebellum **(E)**. GAPDH was used as the loading control for the cell line and mouse and 18S for patient samples (n=3; ns=non-significant, =P>0.05; ***=P<0.005; **=P<0.01; *=P<0.05).

Fig. S9: Purification and 3'-phosphatase activity of Drosophila PNKP and assessment of RNAP II level in SCA3 Drosophila

(A) Affinity purification of 6X-His tagged Drosophila PNKP (CG9601) following IPTG mediated induction (0.5 mM, 16^0C, overnight) (lane 3). Lane 2: Uninduced control, lane 4: purified PNKP after elution from Co^{2+}-resin. Lane 1: Molecular wt. marker.

(B) 3'-phosphatase assay of Drosophila PNKP (lanes 2-3, 10 and 20 ng) vs. human PNKP (lanes 4-5, 10 and 20 ng). Lane 1: No protein, substrate only.

(C) and (D) Upper Panel: Western blots showing the levels of proteins (as indicated in the right of each row) in the whole cell extracts of eye (GMR) **(C)** and larval brain motor neurons (OK6). **(D)**. **Lower Panel:** Quantitation of decrease in RNAP II level (n=2, ***=P<0.005). Actin: used as loading control.

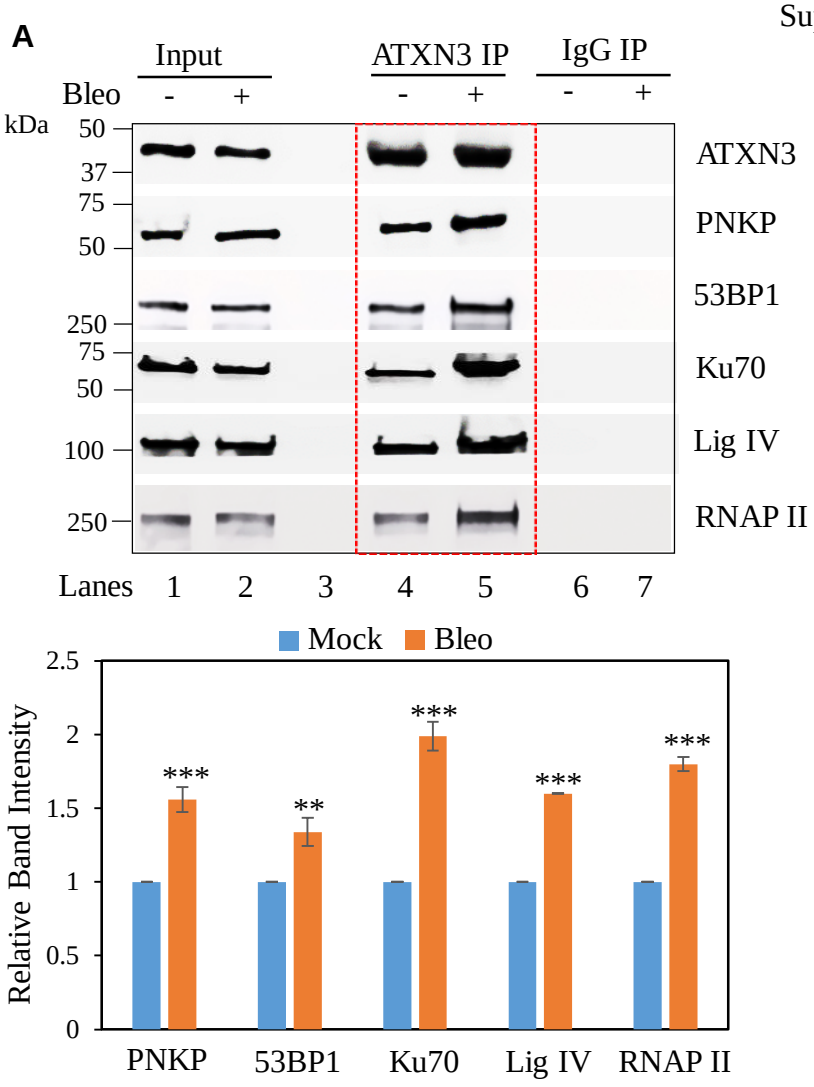
Fig. S10: Immunohistochemistry of Drosophila

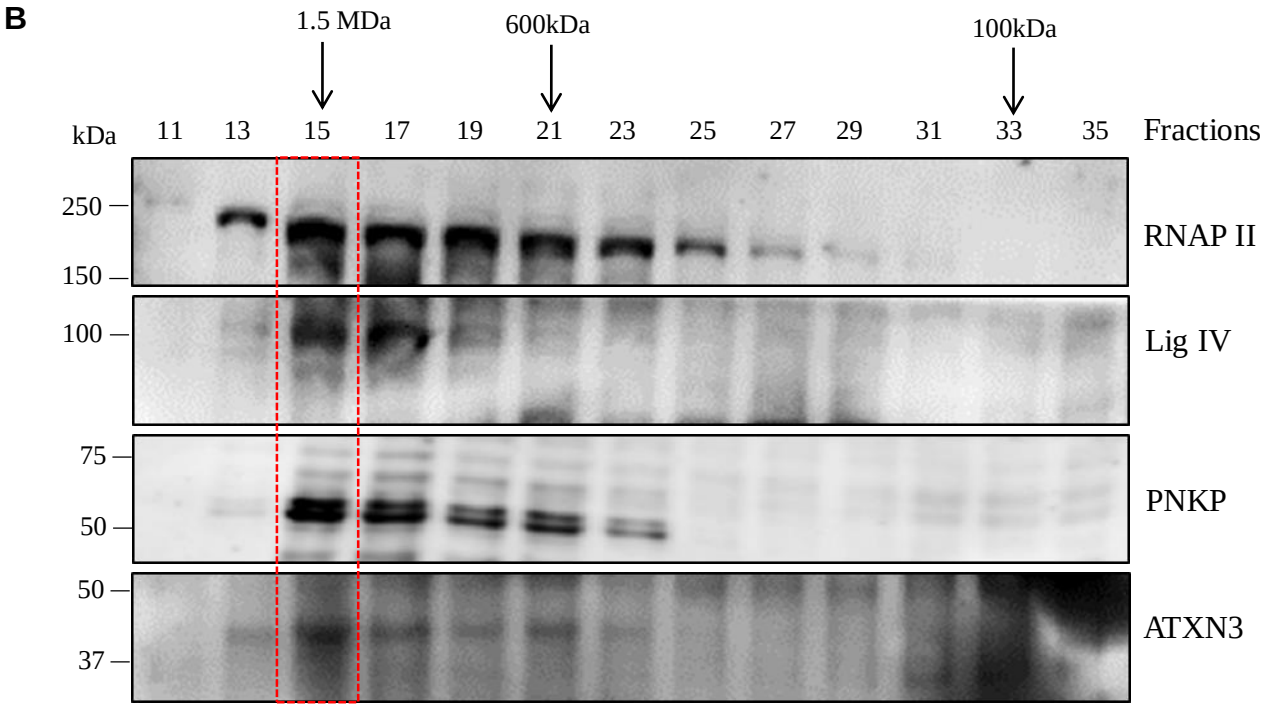
Representative immunofluorescent staining of γ-H2A.v **(Right Panels)** and Dapi **(Left Panels)** in Drosophila larval brain of Control-OK6 **(A)**, with motor neuron specific overexpression of ATXN3.trQ78 alone **(B)** or ATXN3.trQ78 with GS-PNKP **(C)**. Expression of ATXN3.trQ78 leads to elevated number of γ-H2A.v foci and co-expression of PNKP leads to fewer γ-H2A.v foci in the nuclei of motor neurons with ATXN3 expression (as calculated from 6 independent microscopic fields/sample) **(D)**. Error bars represent ±SD of the mean. (n=3, ***=P<0.005). Scale bar: 10 μm.

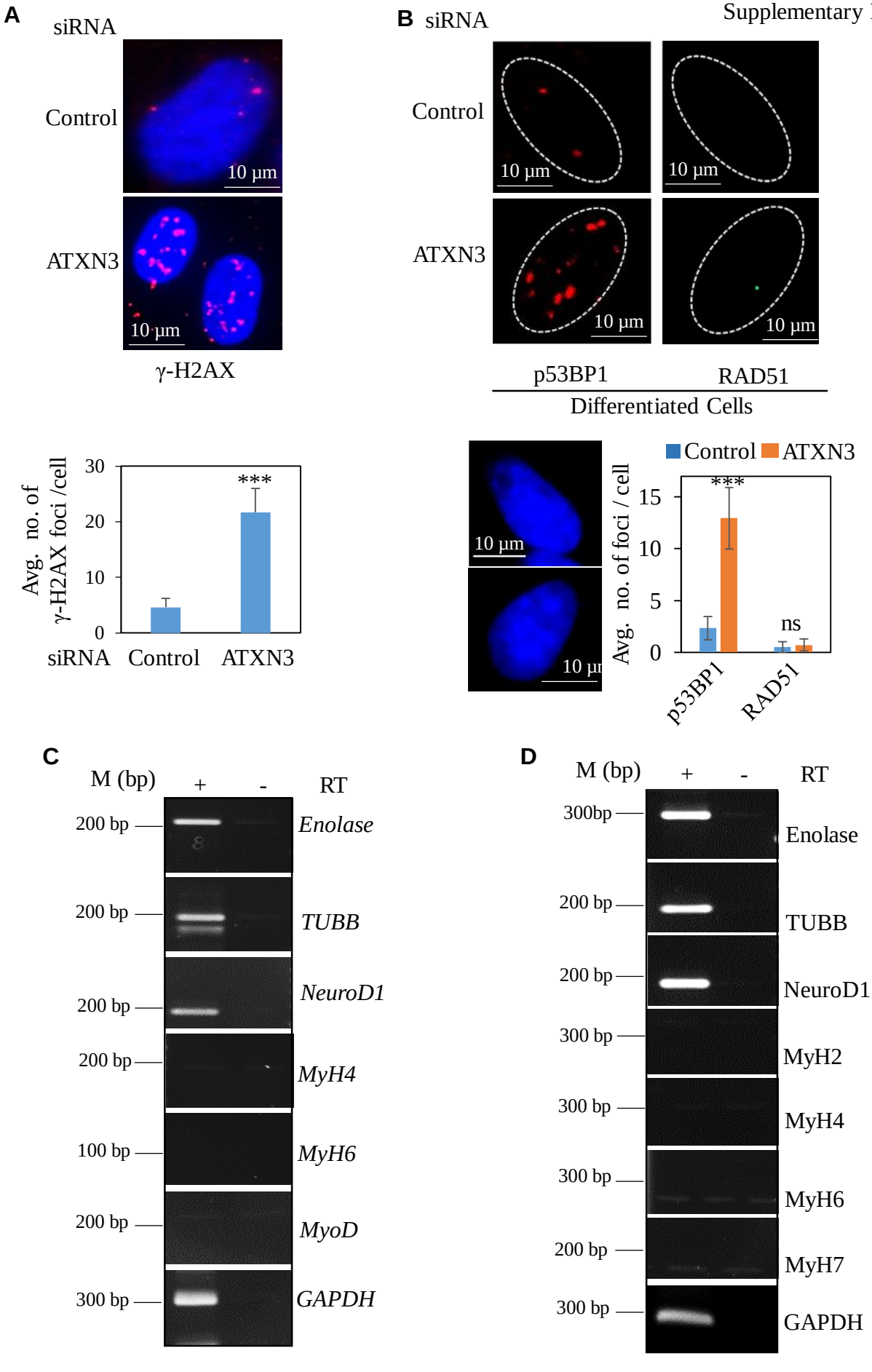
Supplementary references

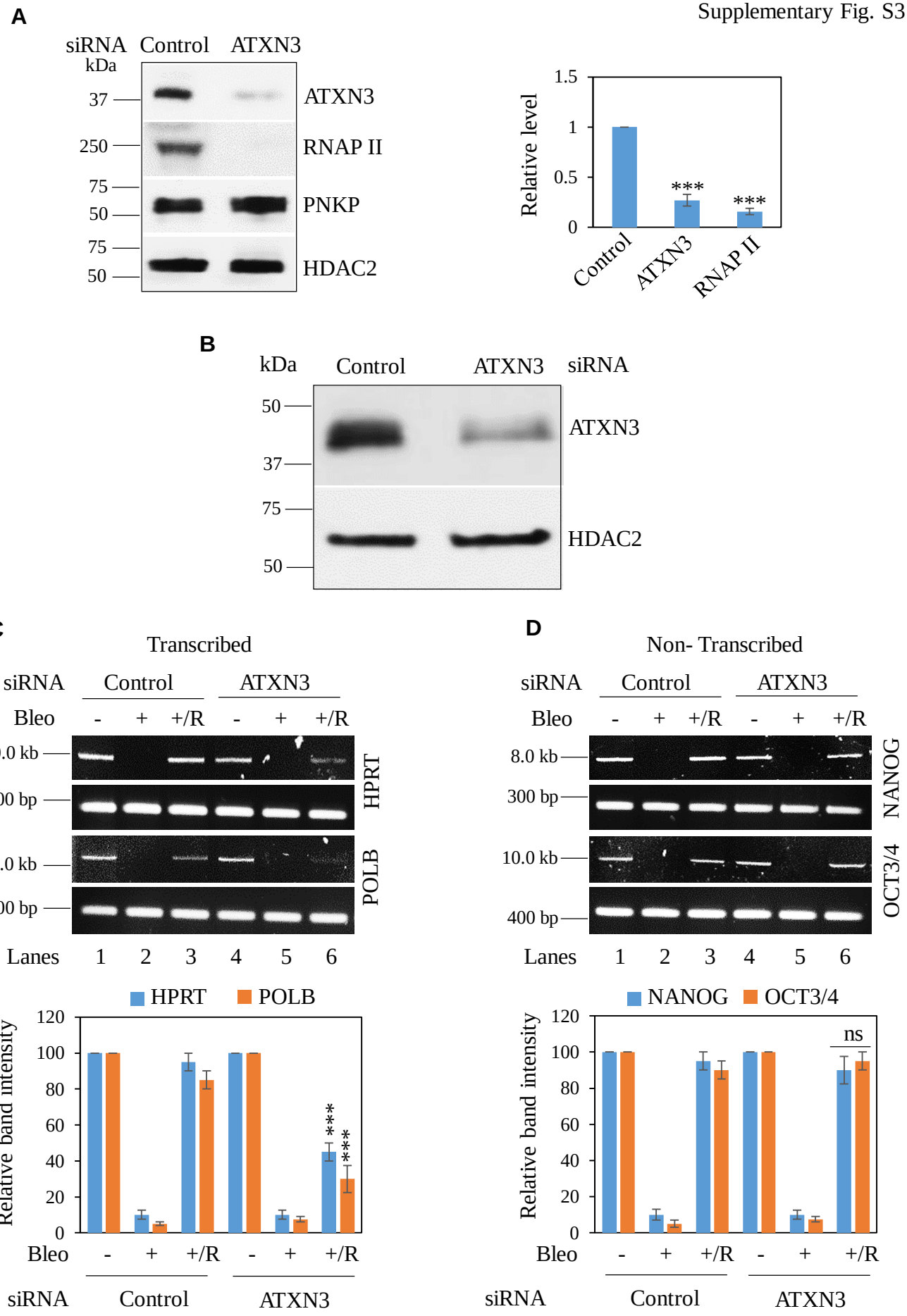
1. A. Chakraborty *et al.*, Classical non-homologous end-joining pathway utilizes nascent RNA for error-free double-strand break repair of transcribed genes. *Nature communications* **7**, 13049 (2016).
2. V. Vasquez *et al.*, Chromatin-Bound Oxidized alpha-Synuclein Causes Strand Breaks in Neuronal Genomes in in vitro Models of Parkinson's Disease. *Journal of Alzheimer's disease : JAD* **60**, S133-S150 (2017).
3. J. Mitra *et al.*, Motor neuron disease-associated loss of nuclear TDP-43 is linked to DNA double-strand break repair defects. *Proceedings of the National Academy of Sciences of the United States of America* 10.1073/pnas.1818415116 (2019).
4. A. Chakraborty *et al.*, *Neil2*-null Mice Accumulate Oxidized DNA Bases in the Transcriptionally Active Sequences of the Genome and Are Susceptible to Innate Inflammation. *The Journal of biological chemistry* **290**, 24636-24648 (2015).
5. J. D. Dignam, R. M. Lebovitz, R. G. Roeder, Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic acids research* **11**, 1475-1489 (1983).
6. L. Wiederhold *et al.*, AP endonuclease-independent DNA base excision repair in human cells. *Molecular cell* **15**, 209-220 (2004).
7. S Kim *et al.*, Telomere dysfunction and cell survival: roles for distinct TIN2-containing complexes *The Journal of Cell Biology* **181(3)**, 447-460 (2008).
8. S. M. Mandal *et al.*, Role of human DNA glycosylase Nei-like 2 (NEIL2) and single strand break repair protein polynucleotide kinase 3'-phosphatase in maintenance of mitochondrial genome. *The Journal of biological chemistry* **287**, 2819-2829 (2012).
9. J. H. Santos, J. N. Meyer, B. S. Mandavilli, B. Van Houten, Quantitative PCR-based measurement of nuclear and mitochondrial DNA damage and repair in mammalian cells. *Methods Mol Biol* **314**, 183-199 (2006).
10. R. Gao *et al.*, Mutant huntingtin impairs PNKP and ATXN3, disrupting DNA repair and transcription. *eLife* **8** (2019).

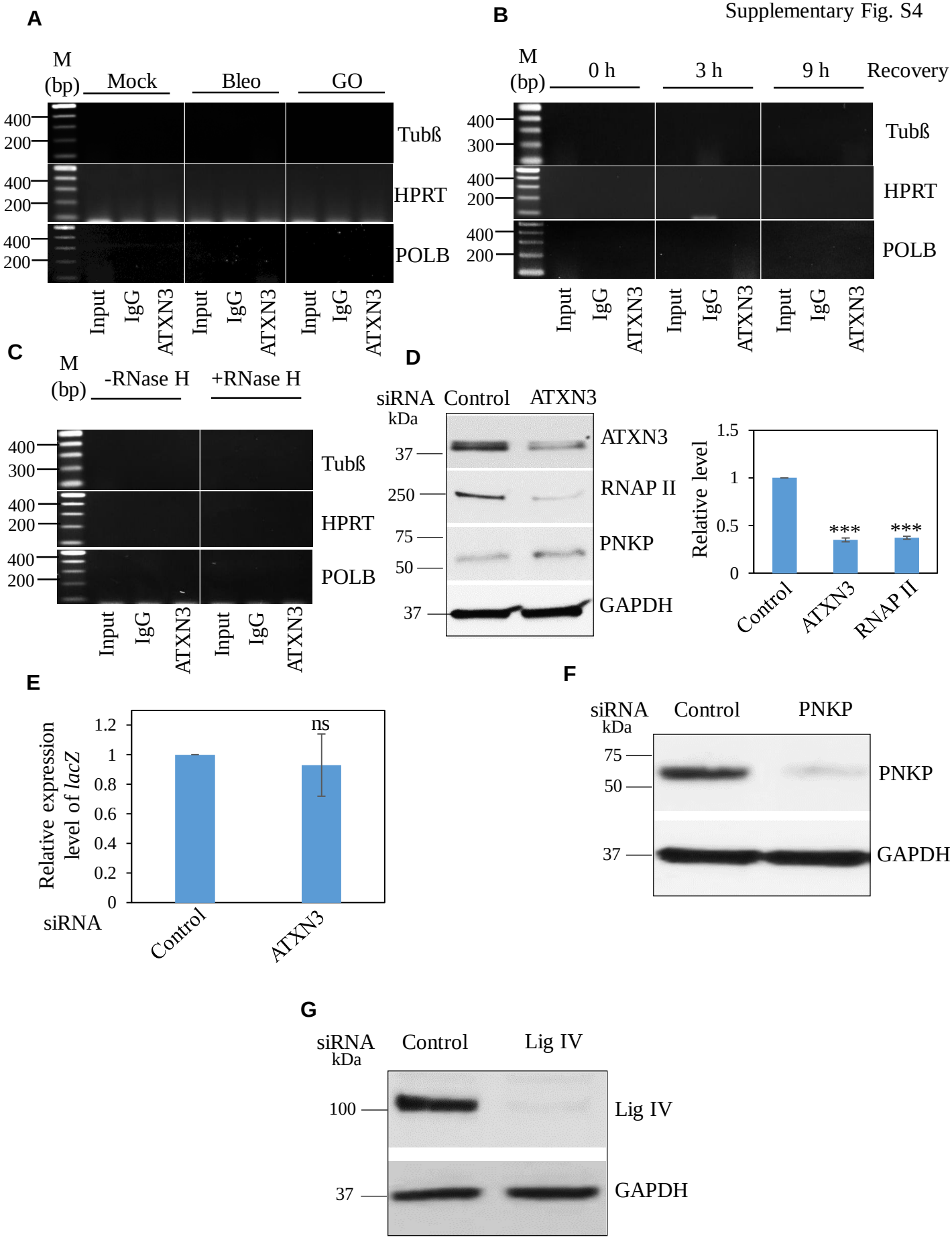
11. A. Silva-Fernandes *et al.*, Chronic treatment with 17-DMAG improves balance and coordination in a new mouse model of Machado-Joseph disease. *Neurotherapeutics* **11**, 433-449 (2014).
12. A. Chatterjee *et al.*, The Role of the Mammalian DNA End-processing Enzyme Polynucleotide Kinase 3'-Phosphatase in Spinocerebellar Ataxia Type 3 Pathogenesis. *PLoS genetics* **11**, e1004749 (2015).
13. H. Dou, S. Mitra, T. K. Hazra, Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *The Journal of biological chemistry* **278**, 49679-49684 (2003).

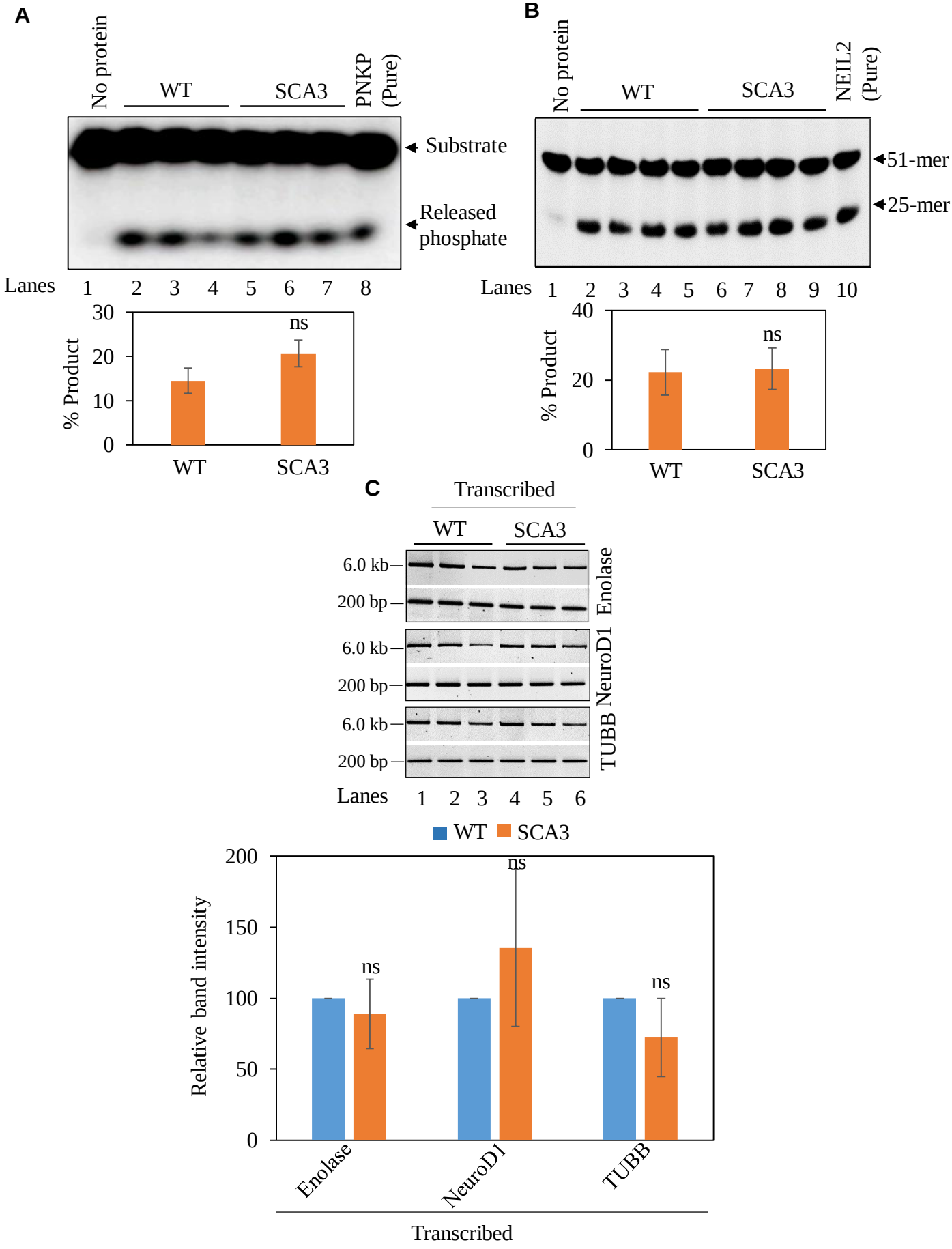

A
Input
ATXN3 IP
IgG IP
Bleo
- + - + - +
kDa
50
37
75
50
250
75
50
100
250
ATXN3
PNKP
53BP1
Ku70
Lig IV
RNAP II
Lanes 1 2 3 4 5 6 7
Mock
Bleo
Relative Band Intensity
2.5
2
1.5
1
0.5
0
***
**
***
***
***
PNKP
53BP1
Ku70
Lig IV
RNAP II

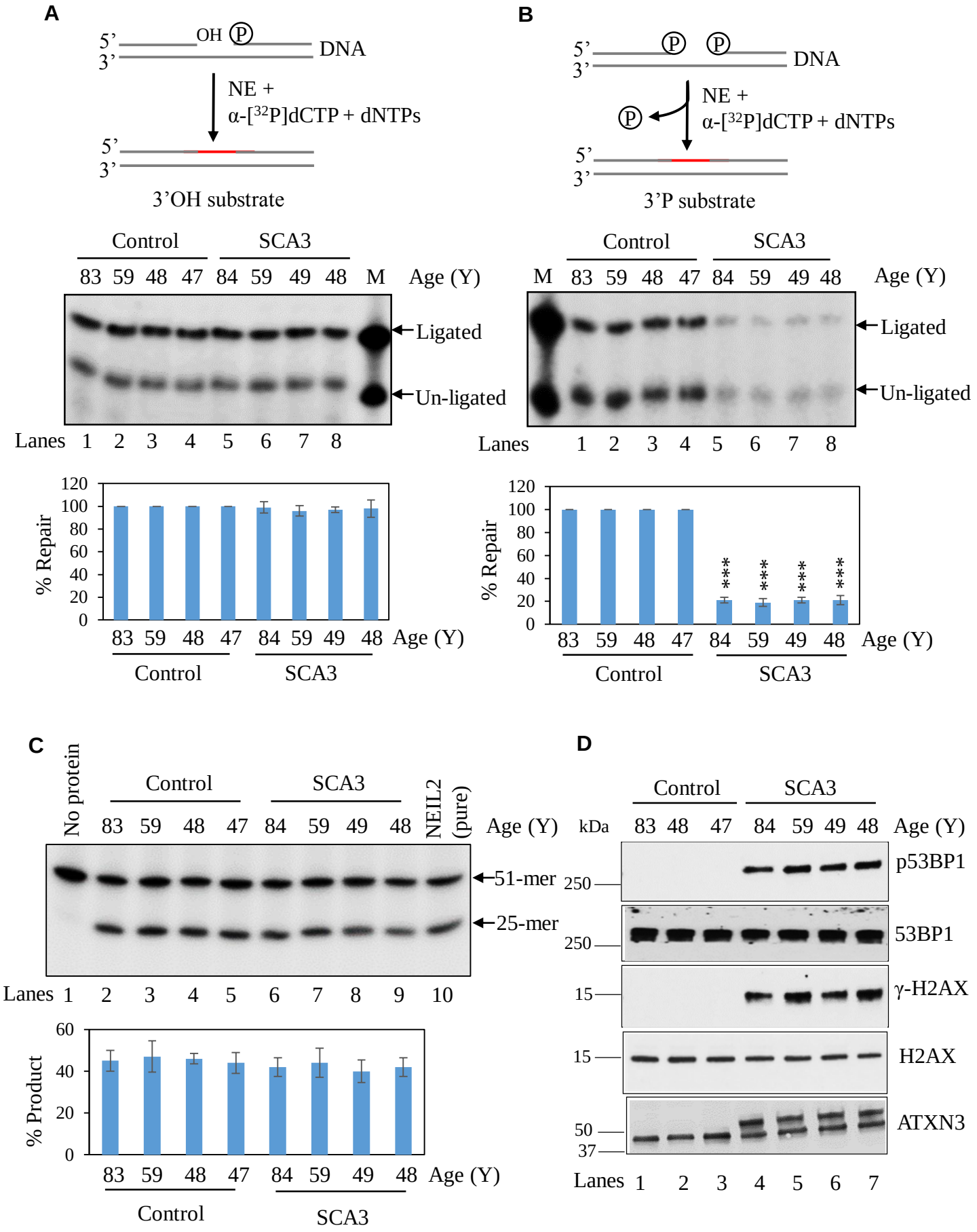

B
1.5 MDa
600kDa
100kDa
kDa
11 13 15 17 19 21 23 25 27 29 31 33 35
Fractions
250
150
100
75
50
50
37
RNAP II
Lig IV
PNKP
ATXN3

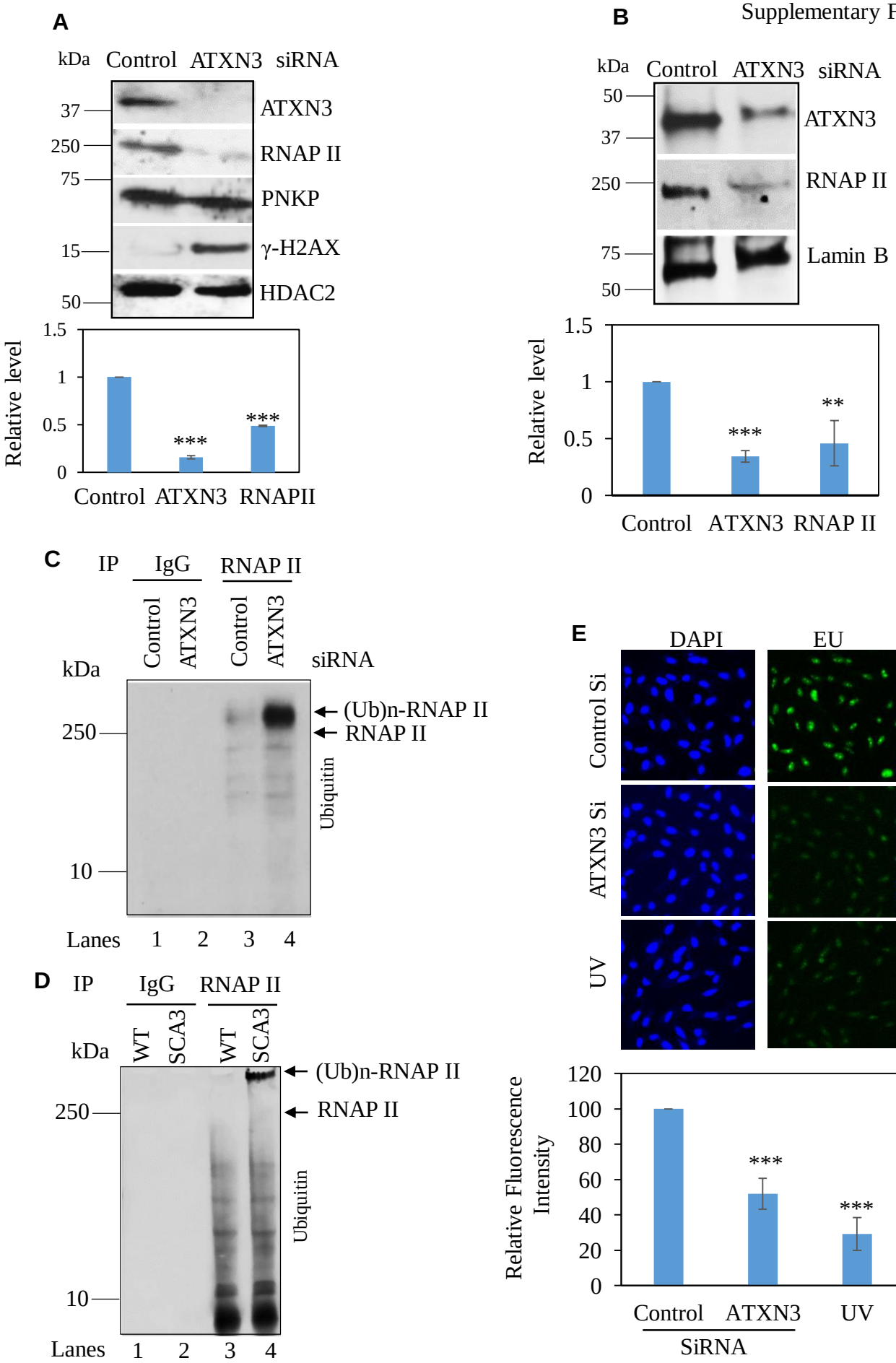

A
siRNA
Control
ATXN3
10 µm
γ-H2AX
Avg. no. of γ-H2AX foci /cell
30
20
10
0
***
siRNA
Control
ATXN3
B
siRNA
Control
ATXN3
10 µm
p53BP1
RAD51
Differentiated Cells
Control
ATXN3
Avg. no. of foci / cell
15
10
5
0
***
ns
p53BP1
RAD51
C
M (bp)
+
-
RT
200 bp
Enolase
200 bp
TUBB
200 bp
NeuroD1
200 bp
MyH4
100 bp
MyH6
200 bp
MyoD
300 bp
GAPDH
D
M (bp)
+
-
RT
300bp
Enolase
200 bp
TUBB
200 bp
NeuroD1
300 bp
MyH2
300 bp
MyH4
300 bp
MyH6
200 bp
MyH7
300 bp
GAPDH

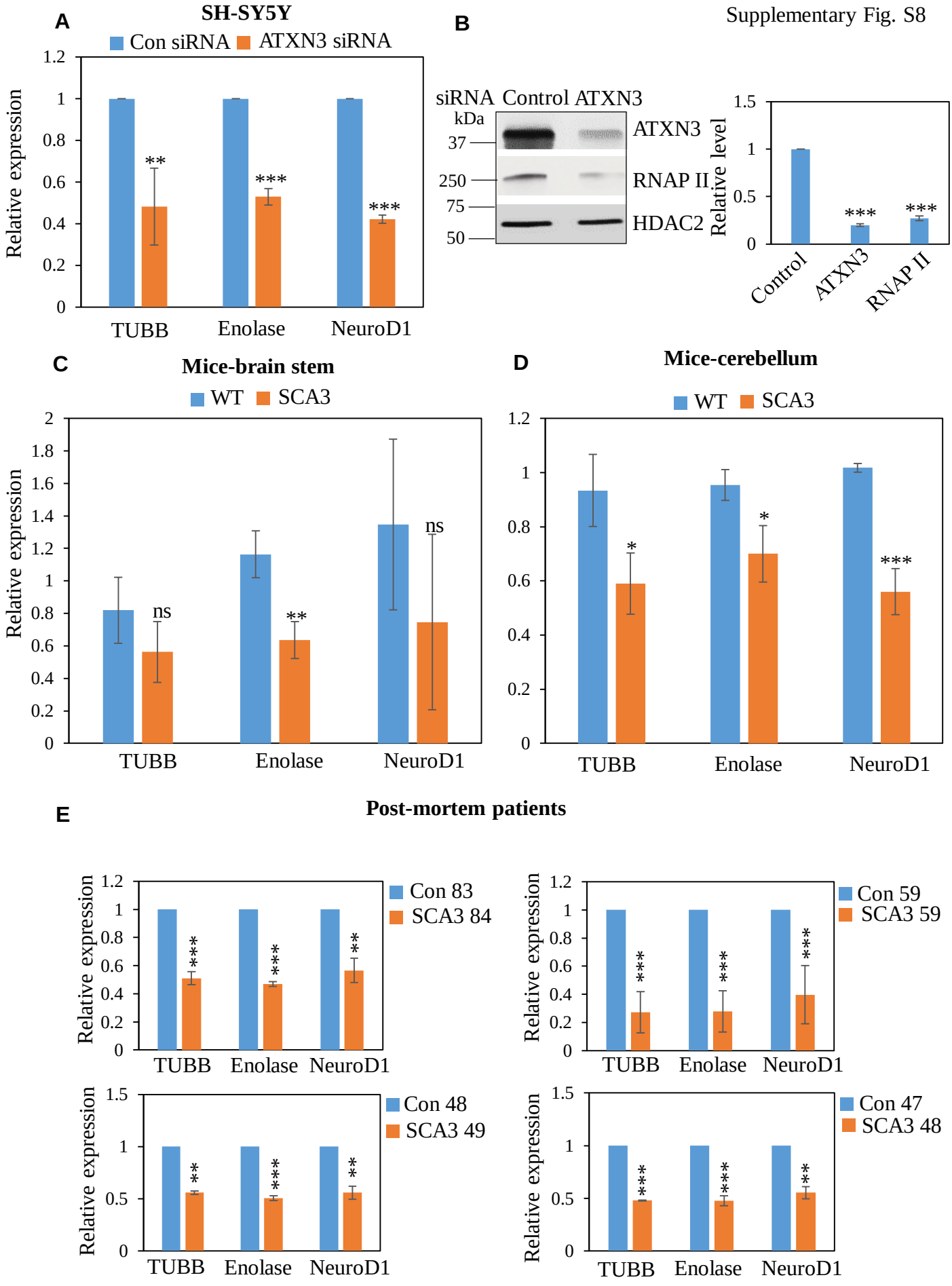

A
siRNA Control ATXN3
kDa
37
250
75
50
75
50
ATXN3
RNAP II
PNKP
HDAC2
Relative level
1.5
1
0.5
0
***
***
Control
ATXN3
RNAP II
B
kDa
Control
ATXN3
siRNA
50
37
75
50
ATXN3
HDAC2
C
Transcribed
siRNA Control ATXN3
Bleo - + +/R - + +/R
10.0 kb
300 bp
10.0 kb
200 bp
HPRT
POLB
Lanes 1 2 3 4 5 6
HPRT POLB
Relative band intensity
120
100
80
60
40
20
0
***
***
Bleo - + +/R - + +/R
siRNA Control ATXN3
D
Non- Transcribed
siRNA Control ATXN3
Bleo - + +/R - + +/R
8.0 kb
300 bp
10.0 kb
400 bp
NANOG
OCT3/4
Lanes 1 2 3 4 5 6
NANOG OCT3/4
Relative band intensity
120
100
80
60
40
20
0
ns
Bleo - + +/R - + +/R
siRNA Control ATXN3

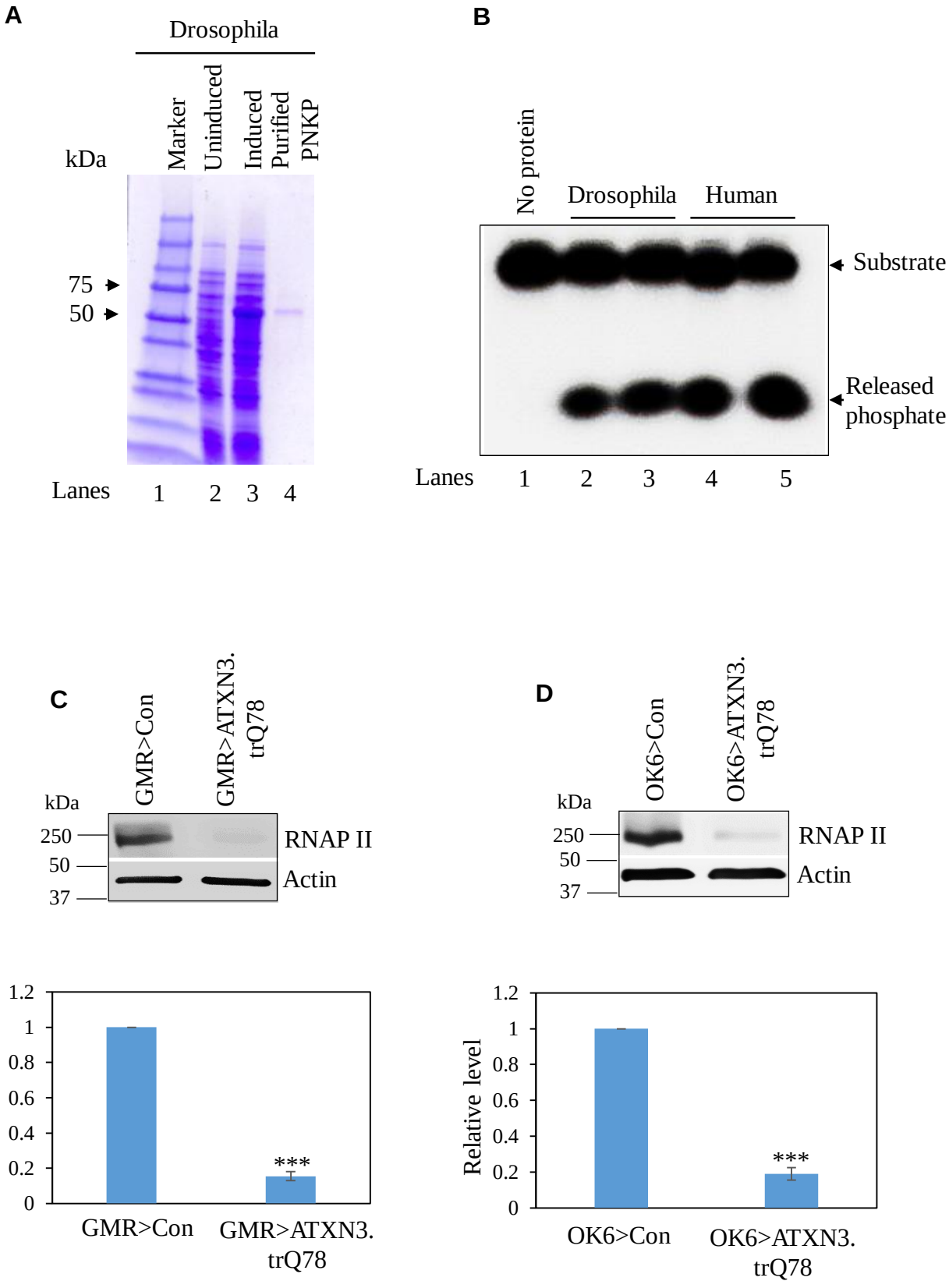

A
M (bp)
Mock
Bleo
GO
400
200
Tubß
HPRT
POLB
Input
IgG
ATXN3
B
M (bp)
0 h
3 h
9 h
Recovery
400
300
200
C
-RNase H
+RNase H
D
siRNA
Control
ATXN3
kDa
37
250
75
50
ATXN3
RNAP II
PNKP
GAPDH
Relative level
1.5
1
0.5
0
***
Control
ATXN3
RNAP II
E
Relative expression level of lacZ
1.2
0.8
0.6
0.4
0.2
ns
siRNA
F
PNKP
G
Lig IV
100

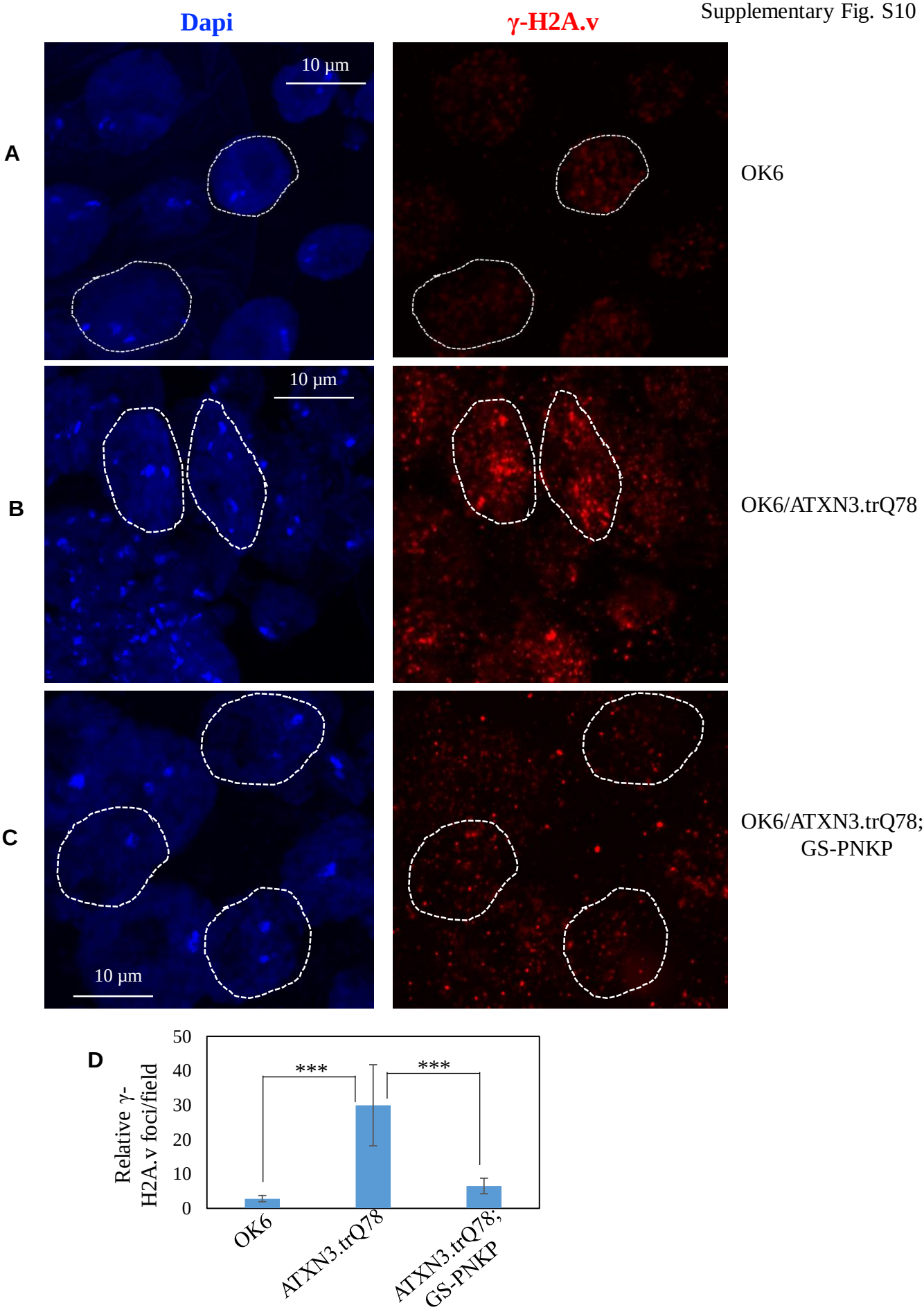

A
No protein
WT
SCA3
PNKP (Pure)
Substrate
Released phosphate
Lanes 1 2 3 4 5 6 7 8
% Product
30
20
10
0
ns
WT
SCA3
B
No protein
WT
SCA3
NEIL2 (Pure)
51-mer
25-mer
Lanes 1 2 3 4 5 6 7 8 9 10
% Product
40
20
0
ns
WT
SCA3
C
Transcribed
WT
SCA3
6.0 kb
200 bp
Enolase
NeuroD1
TUBB
Lanes 1 2 3 4 5 6
WT
SCA3
Relative band intensity
200
150
100
50
0
ns
ns
ns
Enolase
NeuroD1
TUBB
Transcribed


A
5'
3'
OH
P
DNA
NE +
α-[32P]dCTP + dNTPs
3'OH substrate
Control
SCA3
83 59 48 47 84 59 49 48 M
Age (Y)
Ligated
Un-ligated
Lanes 1 2 3 4 5 6 7 8
% Repair
120
100
80
60
40
20
0
83 59 48 47 84 59 49 48 Age (Y)
Control
SCA3
B
5'
3'
P
DNA
NE +
α-[32P]dCTP + dNTPs
3'P substrate
Control
SCA3
M 83 59 48 47 84 59 49 48
Age (Y)
Ligated
Un-ligated
Lanes 1 2 3 4 5 6 7 8
% Repair
120
100
80
60
40
20
0
****
83 59 48 47 84 59 49 48 Age (Y)
Control
SCA3
C
No protein
Control
SCA3
83 59 48 47 84 59 49 48
NEIL2 (pure)
Age (Y)
51-mer
25-mer
Lanes 1 2 3 4 5 6 7 8 9 10
% Product
60
40
20
0
83 59 48 47 84 59 49 48 Age (Y)
Control
SCA3
D
Control
SCA3
kDa
83 48 47 84 59 49 48
Age (Y)
250
p53BP1
250
53BP1
15
γ-H2AX
15
H2AX
50
37
ATXN3
Lanes 1 2 3 4 5 6 7


A
kDa
Control
ATXN3 siRNA
37
250
75
15
50
ATXN3
RNAP II
PNKP
γ-H2AX
HDAC2
Relative level
1.5
1
0.5
0
***
***
Control
ATXN3
RNAPII
B
kDa
Control
ATXN3 siRNA
50
37
250
75
50
ATXN3
RNAP II
Lamin B
Relative level
1.5
1
0.5
0
***
**
Control
ATXN3
RNAP II
C
IP
IgG
RNAP II
Control
ATXN3
Control
ATXN3
siRNA
kDa
250
10
(Ub)n-RNAP II
RNAP II
Ubiquitin
Lanes
1
2
3
4
D
IP
IgG
RNAP II
WT
SCA3
WT
SCA3
kDa
250
10
(Ub)n-RNAP II
RNAP II
Ubiquitin
Lanes
1
2
3
4
E
DAPI
EU
Control Si
ATXN3 Si
UV
Relative Fluorescence Intensity
120
100
80
60
40
20
0
***
***
Control
ATXN3
UV
SiRNA


A
SH-SY5Y
Con siRNA
ATXN3 siRNA
Relative expression
TUBB
Enolase
NeuroD1
B
siRNA Control ATXN3
kDa
37
250
75
50
ATXN3
RNAP II
HDAC2
Relative level
Control
ATXN3
RNAP II
C
Mice-brain stem
WT
SCA3
Relative expression
ns
D
Mice-cerebellum
E
Post-mortem patients
Con 83
SCA3 84
Con 59
SCA3 59
Con 48
SCA3 49
Con 47
SCA3 48


A
Drosophila
Marker
Uninduced
Induced
Purified PNKP
kDa
75
50
Lanes 1 2 3 4
B
No protein
Drosophila
Human
Substrate
Released phosphate
Lanes 1 2 3 4 5
C
GMR>Con
GMR>ATXN3. trQ78
kDa
250
50
37
RNAP II
Actin
Relative level
1.2
1
0.8
0.6
0.4
0.2
0
***
GMR>Con
GMR>ATXN3. trQ78
D
OK6>Con
OK6>ATXN3. trQ78
kDa
250
50
37
RNAP II
Actin
Relative level
1.2
1
0.8
0.6
0.4
0.2
0
***
OK6>Con
OK6>ATXN3. trQ78


Dapi
γ-H2A.v
A
10 µm
OK6
B
10 µm
OK6/ATXN3.trQ78
C
10 µm
OK6/ATXN3.trQ78;
GS-PNKP
D
Relative γ-H2A.v foci/field
50
40
30
20
10
0
***
***
OK6
ATXN3.trQ78
ATXN3.trQ78;
GS-PNKP