PNAS
www.pnas.org

Supplementary Information for

Cytoplasmic synthesis of endogenous *Alu* complementary DNA via reverse transcription and implications in age-related macular degeneration

Shinichi Fukuda[a,b,c,1], Akhil Varshney[a,b,1], Benjamin J. Fowler[d], Shao-bin Wang[a,b], Siddharth Narendran[a,b,e], Kameshwari Ambati[a,b], Tetsuhiro Yasuma[d,f], Joseph Magagnoli[g,h], Hannah Leung[a,b], Shuichiro Hirahara[a,b,i], Yosuke Nagasaka[a,b], Reo Yasuma[a,b,f], Ivana Apicella[a,b], Felipe Pereira[a,b,j], Ryan D. Makin[a,b], Eamonn Magner[k], Xinan Liu[k], Jian Sun[a,b], Mo Wang[l], Kirstie Baker[l], Kenneth M. Marion[l], Xiwen Huang[k], Elmira Baghdasaryan[l,m], Meenakshi Ambati[a,b,n], Vidya L. Ambati[n], Akshat Pandey[a,b], Lekha Pandya[a,b], Tammy Cummings[g,h], Daipayan Banerjee[a,b], Peirong Huang[a,b], Praveen Yerramothu[a,b], Genrich V. Tolstonog[o], Ulrike Held[p], Jennifer A. Erwin[q], Apua C.M. Paquola[q], Joseph R. Herdy[r], Yuichiro Ogura[i], Hiroko Terasaki[f], Tetsuro Oshika[c], Shaban Darwish[s,t], Ramendra K. Singh[s], Saghar Mozaffari[s], Deepak Bhattarai[u], Kyung Bo Kim[u], James W. Hardin[g,v], Charles L. Bennett[g,h,w], David R. Hinton[x], Timothy E. Hanson[y,z], Christian Röver[aa], Keykavous Parang[s], Nagaraj Kerur[a,b,bb,cc], Jinze Liu[k], Brian C. Werner[dd], S. Scott Sutton[g,h], Srinivas R. Sadda[l,m], Gerald G. Schumann[p], Bradley D. Gelfand[a,b,ee], Fred H. Gage[y,2], Jayakrishna Ambati[a,b,cc,ff,2]

Fred H. Gage and Jayakrishna Ambati
Email: ja9qr@virginia.edu, gage@salk.edu

This PDF file includes:

Supplementary text
Figures S1 to S24
Tables S1 to S12
SI References

Supplementary Information Text

Expanded manuscript methods

Animals

Wild-type (WT) C57BL/6J mice were purchased from The Jackson Laboratory (Bar Harbor, ME) and Envigo (Frederick, MD), respectively. *Casp1*$^{-/-}$ *Casp4*$^{129mt/129mt}$ (denoted Casp1/4 dko) mice were obtained from G. Nuñez (University of Michigan School of Medicine). For all procedures, anesthesia was achieved by intraperitoneal injection of 100 mg/kg ketamine hydrochloride (Ft. Dodge Animal Health) and 10 mg/kg xylazine (Phoenix Scientific), and pupils were dilated with topical 1% tropicamide and 2.5% phenylephrine hydrochloride (Alcon Laboratories). Mice were treated in accordance with the guidelines of the University of Virginia and University of Kentucky Institutional Animal Care and Use Committees and the Association for Research in Vision and Ophthalmology. Both male and female mice between 6–10 weeks of age were used.

Fundus photography

Fundus imaging of dilated mouse and rat eyes was performed using a TRC-50 IX camera (Topcon) linked to a digital imaging system (Sony).

Chemicals

The NRTIs lamivudine (3TC) was purchased from Sigma-Aldrich. Azidothymidine triphosphate (AZT-TP) was purchased from TriLink Biotechnologies. A diethyl-modified version of AZT (DE-AZT) were synthesized as previously described (1). The nuclear-targeting cyclic peptides (Cpep) and Cpep-conjugated 3TC (Cpep-3TC), and cytoplasmic-targeting PA-4 3TC have been previously described (2, 3).

siRNA sequences and treatment

Two siRNAs (mL1 siRNA 1 and mL1 siRNA 2) targeting three mouse L1 subfamilies (T_F, G_F, A) and two siRNAs (hL1 siRNA 1 and hL1 siRNA 2) targeting human L1-Hs were designed and purchased from Dharmacon. For in vivo animal use, the oligonucleotides contained 3′ cholesterol moieties to enable cell penetration. Their sequences and the maps of their targeted L1 locations are presented in fig. S1A and fig. S10. siRNAs were formulated in siRNA buffer (20 mM KCl, 0.2 mM $MgCl_2$ in HEPES buffer at pH 7.5; Dharmacon) or phosphate-buffered saline (PBS; Sigma-Aldrich). hL1 siRNA and mL1 siRNA were transfected into human and mouse RPE cells using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. 1 µl of cholesterol-conjugated siRNAs (2 µg/µl) targeting mouse L1 or *Luc* (luciferase control) (both from Dharmacon) were subretinally injected two days before *Alu* RNA or vehicle administration.

In vitro transcribed *Alu* RNA and *Alu* mutant RNA

Alu RNA was synthesized from a linearized plasmid containing a consensus *Alu*Y sequence with an adjacent 5′ T7 promoter (4), subjected to AmpliScribe T7 Flash Transcription kit (Epicentre) according to the manufacturer's instructions. DNase-treated RNA was purified using MEGAclear (Ambion) and integrity confirmed by agarose gel electrophoresis. *Alu* Y RNA with G25C/G159C mutations was synthesized from a linearized plasmid containing the G25C/G159C mutations as described above.

Assessment of RPE degeneration

Subretinal injections were performed as previously described (4–9). Seven days after subretinal injection, RPE health was assessed by fundus photography and immunofluorescence staining of zonula occludens-1 (ZO-1) on RPE flat mounts (whole mount of posterior eye cup containing RPE and choroid layers). Mouse RPE/choroid flat mounts were fixed with 2% paraformaldehyde, stained with rabbit polyclonal antibodies against mouse ZO-1 (1:100, Invitrogen) and visualized with Alexa-594 (Invitrogen). All images were obtained by microscopy (model SP-5, Leica or A1R Nikon confocal microscope system, Nikon). Imaging was performed by an operator blinded to the group assignments.

Quantification of RPE degeneration

Quantification of RPE degeneration was performed using two methodologies (binary assignment and cellular morphometry) as described previously (1): Binary assignment (healthy versus unhealthy) (6–9) was independently performed by two blinded raters (inter-rater agreement = 99.5%; Pearson $r^2 = 0.98$, $P < 0.0001$; Fleiss $\kappa = 0.99$, $P < 0.0001$). Quantifying cellular morphometry for hexagonally packed cells was performed in semi-automated fashion by three masked graders by adapting our previous analysis of the planar architecture of corneal endothelial cell density (10), which resembles the RPE in its polygonal tessellation. We quantified polymegethism (coefficient of variation of cell size), a prominent geometric feature of RPE cells in GA (5, 11–14), using the Konan Cell Check software (Ver. 4.0.1), a commercial U.S. FDA-cleared software that has been used for registration clinical trials, as previously described (10).

Subretinal and intravitreous injections

Subretinal injections (1 µl) or intravitreous injections (0.5 µl) in mice or rat were performed using a 35-gauge needle (Ito Co. Fuji, Japan). In vivo transfection of *Alu* RNA or mutant *Alu* RNA (300 ng per eye) as previously described (4, 5). 3TC, Cpep-3TC, Cpep, PA-4 3TC, or olaparib (25 ng) (Cayman Chemical) were administered by intravitreous injection. 1 µl of cholesterol-conjugated siRNAs (2 µg/µl) targeting mouse L1 or *Luc* (luciferase control) (Dharmacon) were subretinally injected three days prior to *Alu* RNA or vehicle administration. For in vivo functional rescue of L1 using mL1 siRNA-refractory plasmids encoding a synthetic, codon-optimized full-length human L1 element (pORFeus-Hs) or rat L1 ORF2p (pORF2-Rn) were used. A total volume of 1 µl of pLD401 (Tet promoter, [*ORFeus*-Hs] full L1 coding sequence, ORF2p-3×FLAG; 0.25µg; gift of J.D. Boeke, NYU Langone Health) (15) was subretinally injected using 10% Neuroporter (Genlantis) along with prtTA (0.25 µg; Addgene) and mL1 siRNA (2µg) and intravitreous injection of doxycycline (4 µg; Sigma-Aldrich) or PBS. A total volume of 1 µl of pORF2-Rn (0.25 µg) (16) and mL1 siRNA (2 µg) was subretinally injected using 10% Neuroporter (Genlantis). Two days later, *Alu* RNA was injected through the same subretinal injection site.

Immunoblotting

Cells and tissue were homogenized in RIPA buffer (Sigma-Aldrich) with protease and phosphatase inhibitors (Roche) or lysed directly in Laemmli buffer. Protein concentration was determined using Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). Equal quantities of protein (10–50 µg) prepared in Laemmli buffer were resolved by SDS-PAGE on Novex® Tris-Glycine Gels (Invitrogen) or Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad) and transferred onto Immobilon-FL PVDF membranes (0.2 or 0.45 µm) (Millipore). The transferred membranes were blocked in Odyssey® Blocking Buffer (PBS) for 1 h at room temperature and then incubated with primary antibody at 4°C overnight. Immunoreactive bands were visualized using species-specific secondary antibodies conjugated with IRDye®. The blot images were captured on Odyssey® imaging systems. Rabbit polyclonal anti-mouse L1 ORF2p (1.0 µg/ml, Rockland Cat#600-401-GT3), rabbit polyclonal anti-human vinculin (1:2,000, Sigma-Aldrich Cat#V4139), chicken polyclonal anti-tubulin (1:50,000, Abcam Cat#ab89984) or mouse monoclonal anti-tubulin (1:5,000, Sigma-Aldrich Cat#T6199) were used.

Cell culture and transfection

Primary mouse and human RPE cells were isolated by adapting previously described protocols (9). All cells were maintained at 37°C in a 5% CO_2 environment. Mouse RPE cells were cultured in DMEM supplemented with 20% FBS and penicillin/streptomycin antibiotics at standard concentrations; primary human RPE cells were maintained in DMEM supplemented with 10% FBS and antibiotics. The human RPE cell line ARPE19 was cultured as previously described (9) and maintained in DMEM-F12 containing penicillin/streptomycin, Fungizone, and gentamicin. HEK293T cells were cultured in DMEM with 10% fetal bovine serum (FBS) with 100 U/ml penicillin/streptomycin and 2 mM L-glutamine. Primary human subcutaneous pre-adipocytes (ATCC, PCS-210-010) and primary human dermal fibroblasts (ATCC, PCS-201-012) were grown in fibroblast basal medium with fibroblast growth kit for low serum (ATCC). Umbilical artery vascular smooth muscle cells (Lonza, CC-2579) were grown in SmGM-2 (Smooth Muscle Growth Medium-2) BulletKit (Lonza). Primary human skeletal myoblasts (GIBCO, A11440) were grown in

DMEM with 2% horse serum. Primary human epidermal keratinocytes (Thermo Fisher Scientific, C0215C) were grown in EpiLife Medium (GIBCO). Human umbilical vein endothelial cells (HUVEC) were grown in HUVEC EGM™-2 Media (Lonza). Primary human peripheral blood mononuclear cells (ATCC, PCS-800-011) were directly used without culture for experiments. RPMI 1640 medium with 10% human serum with human GM-CSF (Miltenyi Biotec) was used as media during the experiment. Transfections were performed according to the manufacturer's instructions (Lipofectamine 2000, Invitrogen). For transfection of primary human peripheral blood mononuclear cells, HiPerFect transfectant (Qiagen) was used as previously described (17). NRTIs were administered 60 min before transfection and added again upon replacement of media at 8 h. To induce acid injury or osmotic stress, primary human RPE cells were treated with HCl (pH 4.0 media) or H_2O for 30, 1h and 2h at 37°C in 5% CO_2, respectively. NIH3T3 Tet-ON cells were cultured in DMEM with 10% tetracycline-free fetal bovine serum (FBS) with 100 U/ml penicillin/streptomycin.

Induction of *Alu* RNA by various stimuli

In vitro transcribed *Alu* Y RNA, *DICER1* antisense (AS) oligonucleotide (5′-GCUGACCTTTTTGCTUCUCA-3′), or control scrambled AS (5′-TTGGTACGCATACGTGTTGACTGTGA-3′) (Integrated DNA Technologies) were transfected into human and mouse RPE cells using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Heat shock was induced by placing cells in a 42°C incubator for 20 min and then allowed to recover at 37°C for 1 h (18).

In situ hybridization

RPE from mice was collected at 24 h after subretinal injection. Cells in culture were collected after 6–8 h after *Alu* RNA transfection. RPE flat mounts or cells were fixed in 4% PFA/PBS for 20 min. For *Alu* cDNA detection, all samples were treated with RNase A. To confirm whether the target was single-stranded DNA, S1 nuclease (Thermo Fisher Scientific) was treated for 30 min at room temperature. RNA probes, prepared from linearized *Alu* cDNA templates using a T7 fluorescein RNA labeling kit was hybridized overnight at 37°C in a mixture containing 10% dextran sulfate, 2 mM vanadyl-ribonucleoside complex, 0.02% RNase-free BSA, 40 μg *E. coli* tRNA, 2× SSC, 50% formamide, and RNA probe. Cells were then subjected twice to stringent washing at 50°C in 50% formamide, 0.1× SSC for 30 min. Following washing, samples were incubated with a horseradish peroxidase (HRP)-conjugated anti-fluorescent antibody (PerkinElmer Cat# NEF710001EA) at a 1:200 dilution. Visualization of fluorescein-labeled probe was performed with the TSA™ plus fluorescence system. The fluorescent was detected using a Leica SP-5 or A1R Nikon confocal microscope system. For L1 overexpression, ARPE-19 cells were transfected with the L1 expression vector pES2TE1 (gift of J.V. Moran, University of Michigan Medical School) (19). Then, *Alu* cDNA synthesis was monitored by in situ hybridization after *Alu* RNA transfection.

Equator blotting

We term "equator blot" as a combination of classic Southern and northern blotting procedures. An equator blot is similar to a Southern blot in that it probes for a target DNA sequence, yet unlike a typical Southern blot, it does not involve restriction enzyme digest of the DNA. Instead, the DNA is run without enzyme digestion prior to hybridization, per the typical northern blot procedure. Hence, we refer the procedure of hybridization of undigested DNA as an equator blot. Total nucleic acid or nuclear and cytoplasmic fractions were extracted from cells as described below. Primary human RPE cells were collected after *Alu*Y RNA transfection and the cytoplasmic fraction was treated with RNase A (Invitrogen). To confirm whether the target was single-stranded DNA or double-strand DNA, S1 nuclease (Thermo Fisher Scientific; 30 min at room temperature) or double-strand DNase (Thermo Fisher Scientific; 2 min at 37°C) were added after RNase treatment, according to the manufacturer's instructions. For human tissue, DNA and RNA were extracted using DNA and RNA Purification Kit (Epicentre); RNase A was added for DNA isolation. DNA samples were run on 10% TBE-urea gels (BioRad) according to the manufacturer's instructions. Samples were transferred and UV crosslinked to a HyBond N+ nylon membrane and

blotted for *Alu* cDNA. *Alu* cDNA biotinylated oligonucleotide probe was synthesized by PCR from a linearized plasmid containing a consensus *Alu*Y element as above using the following primers: for *Alu* cDNA detection (forward 5′-biotin-GGGCCGGGCGCGGTG-3′ and reverse 5′-GTACCTTTAAAGAGACAGAGTCTCGC-3′), and then purified. Blots were developed with the Pierce chemiluminescent nucleic acid detection kit (Thermo Fisher Scientific). The blot images were captured on Odyssey® imaging systems.

Northern blotting

Total nucleic acid or nuclear and cytoplasmic fractions were extracted from cells as described below. For human tissue, DNA and RNA were extracted using DNA and RNA Purification Kit (Epicentre); DNase I was added for RNA isolation. RNA samples were run on 10% TBE-urea gels (BioRad) according to the manufacturer's instructions. Samples were transferred and UV crosslinked to a HyBond N+ nylon membrane and blotted for *Alu* RNA, and U6 RNA. U6 biotinylated oligonucleotide probe was synthesized by Integrated DNA Technologies (5′-CACGAATTTGCGTGTCATCCTT-biotin-3′). *Alu* RNA biotinylated oligonucleotide probe was synthesized by PCR from a linearized plasmid containing a consensus *Alu*Y element as above using the following primers: for *Alu* RNA detection (forward 5′-GGGCCGGGCGCGGTG-3′ and reverse 5′-biotin-GTACCTTTAAAGAGACAGAGTCTCGC-3′) and then purified. Blots were developed with the Pierce chemiluminescent nucleic acid detection kit (Thermo Fisher Scientific). The blot images were captured on Odyssey® imaging systems.

Nuclear and cytoplasmic fractionation

Briefly, cells were collected and lysed with gentle extraction buffer prepared in 1× PBS containing 1% v/v Triton X-100 (Sigma-Aldrich) and 1 mM EDTA for 15 min on ice. Lysates were vortexed and centrifuged at 1,000 × g for 10 min at 4°C. For cytoplasmic fractionation, the supernatant was collected, subjected to repeated centrifugation four times, and then purified using a DNA purification column (Enzymax). Lysis buffer was added to the pellet for reconstitution. Lysate supernatant was vortexed and further centrifuged at 13,000 × g for 2 min at room temperature. Lysate supernatant was used as the nuclear fraction and purified using a DNA purification column (Enzymax). For cDNA detection, samples were treated with RNase A (Ambion) for 30 min at 37°C. To confirm nuclear and cytoplasmic fractionation, cytoplasmic and nuclear RNA were isolated from primary human RPE cells and run on a 0.9% agarose gel to assess genomic DNA, 18S rRNA, and 28S rRNA. Levels of cytoplasmic and nuclear U6 RNA and tRNA were also measured by PCR. PCR reactions were performed using the following primers: *U6* (forward 5′-GTGCTCGCTTCGGCAGCACATATAC-3′; reverse 5′-AAAAATATGGAA CGCTTCACGAATTTG-3′); tRNA (forward 5′-AGCAGAGTGGCGCAGCGG-3′; reverse 5′-GATCCATCGACCTCTGGGTTA-3′). A primer set within an intron of GPR15 to measure genomic DNA was as previously described (20, 21). Cytoplasmic and nuclear levels of GPR15 were directly amplified by real-time PCR (without RT) using the following primers: *GPR15* (forward 5′-GGTCCCTGGTGGCCTTAATT-3′; reverse 5′-TTGCTGGTAATGGGC ACACA-3′).

Alu cDNA detection by real-time PCR

Cells were collected after counting the cell number and the cytoplasmic fraction was treated with RNase A (Ambion). The RNase-treated cytoplasmic fraction was purified with PCR clean-up kit (QIAquick, Qiagen). Then samples were directly amplified by real-time quantitative PCR (Applied Biosystems 7900 HT Fast Real-Time PCR system) with Power SYBR green Master Mix. Primers were specific for human *Alu* cDNA (forward 5′-TTAGCCGGGAGTGGTGTCGG-3′ and reverse 5′-ACCTCCCGGGTTCACGCCATT-3′). The copy number of *Alu* cDNA was calculated using standard curves that were obtained using serial dilutions of the plasmids containing an *Alu*Y sequence. *Alu* cDNA copy number was normalized to cell number.

We developed a method to purify and amplify the reverse transcribed single-stranded DNA and minimize the amplification from contaminating genomic DNA (*Alu* c-PCR). First, total cell lysate was fractionated to yield nuclear and cytoplasmic fractions as above. The purified cytoplasmic fraction was poly-A-tailed with TdT (NEB) for 30 min at 37°C according to the manufacturer's instructions. The poly-A-tailed template was annealed and extended by a PolyT-anchored

adapter primer (TAV oligo). These anchored DNAs were amplified using anchor-specific primer and reverse primer specific for *Alu*. The TAV oligo contains a unique 22-nt anchor sequence at the 5′-end followed by 18 thymidines (dT) and ends with V nucleotide (where V represent adenosine (A), guanosine (G) or cytidine (C)). TAV (5′-GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTTV-3′), anchor primer (5′-GACCACGCGTATCGATGTCGAC-3′) and *Alu* specific primer (5′-ACCTCCCGGGTTCACGCCATT-3′). The *Alu* c-PCR method specifically detects linear *Alu* cDNA while avoiding detecting the circular form of extrachromosomal *Alu* DNAs. In this method, the first step is poly-A-tailing of linear *Alu* cDNAs; this poly A-tailed DNA primes the synthesis of DNA by poly T-anchored adapter primer. These anchored DNAs are then amplified by using a primer specific for the adapter and another primer specific for *Alu*. Circular *Alu* dsDNAs may already have a poly A region that can prime the synthesis of DNA by the poly T-anchored primer; however, this anchored DNA cannot be amplified by using the primer specific for *Alu*.

Real-time PCR

For human tissue, total RNA was extracted using MasterPure™ Complete DNA and RNA Purification Kit (Epicentre) according to the manufacturer's recommendation. We exhaustively digested RNA samples with RNase free DNase (Turbo DNase) before cDNA preparation for L1 expression quantification. Effectiveness of the DNase digestion was assessed using controls that omitted the RT enzyme. The RT products (cDNA) were amplified by real-time quantitative PCR (Applied Biosystems 7900 HT Fast Real-Time PCR system) with Power SYBR green Master Mix. Relative gene expression was determined by $2^{-\Delta\Delta Ct}$ method using 18S rRNA as an internal control. Primers specific for hZNF66 (forward 5′-GCTCCTCTAACCTTACTAAACAC-3′ and reverse 5′-TTTGCCACATTTATTGCACT-3′), hZFP30 (forward 5′-ATAGAAGCCTTTCATCACCT-3′ and reverse 5′-TTGCCCTGAAATACAGTTCC-3′), hGBA2 (forward 5′-CCCAAAAGAGACGGACTGCT-3′ and reverse 5′-AGCCCATGCCTATATGCTT-3′), hLINC01873 (forward 5′-ACGGGAGGACATTCAAACCAA-3′ and reverse 5′- ATCTTCCATCGCTGATACCCT-3′), mZfp933 (forward 5′-ACAGCATAGTAATCTCCGAA-3′ and reverse 5′-AAGATGATAGTAACGTGCAA-3′), mHfe2 (forward 5′-GCCAACGCTACCACCATCCG-3′ and reverse 5′-ACGTGACTCCCAAGGTTAGCA-3′), mZfp945 (forward 5′-GGCTCATATCTTAGAATGCAC-3′ and reverse 5′-GATCTGTCGCAATTACCAC-3′), mPias4 (forward 5′-AGCTTCCGAGTATCAGACCT-3′ and reverse 5′-TGCACTCTTCTTGGCATAGCG-3′) were used.

Ex vivo reverse transcriptase (RT) activity assay

Ex vivo reverse transcriptase (RT) activity in nuclear and cytoplasmic protein fractions was assessed using an *Alu* RNA-templated reaction. Nuclear and cytoplasmic fractionation was performed using NE-PER nuclear cytoplasmic extraction kit (Thermo), as per the manufacturer's instructions. Briefly, in this assay, exogenous *Alu* RNA and *Alu*-R primer (5' ACCTCCCGGGTTCACGCCATT 3') were incubated with nuclear or cytoplasmic protein lysate containing endogenous L1 ORF2p, which acts on *Alu* RNA in trans and gives rise to an *Alu* complementary DNA. The RT reaction was carried out in a 20 µl reaction mix containing *Alu* RNA template (10 ng); *Alu* primer (10 pmol); dNTPs mix; cytoplasmic or nuclear protein, and Quantiscript RT Buffer (Qiagen). The reaction mixture was incubated at 42°C for 30 min. The resulting cDNA was quantified by qPCR using *Alu* RNA template-specific primers. Heat denaturation of nuclear or cytoplasmic fraction was performed by heat inactivation at 95°C for 10 min.

In vitro reverse transcriptase (RT) activity

In vitro reverse transcriptase (RT) activity in nuclear and cytoplasmic protein fractions was assessed using an *Alu* RNA-templated reaction. The RT reaction was carried out in a 20 µl reaction mix containing *Alu* RNA template (10 ng); *Alu* primer (10 pmol); dNTPs mix; cytoplasmic or nuclear protein, and Quantiscript RT Buffer (Qiagen). The reaction mixture was incubated at 42°C for 30 min. The resulting cDNA was quantified by qPCR using *Alu* RNA template-specific primers. The reaction to evaluate self-priming activity of *Alu* RNA was carried out in the absence of priming oligos in a 20 µl reaction mix containing: *Alu* RNA with 3′-U tail; dNTP mix; cytoplasmic

protein from mouse RPE cells; and Quantiscript RT Buffer (Qiagen) as described above. The resulting cDNA product was quantified by qPCR using *Alu* RNA template-specific primers. *Alu* RNA tailed on the 3′ end with chain terminator dideoxy thymidine base (ddTTP) was generated using TdT (NEB) according to the manufacturer's instructions. *Alu* RNA tailed on the 3′ end with chain terminator cordycepin tri-phosphate was generated using PAP (NEB) according to the manufacturer's instructions. An *Alu* RNA transcript embedded with a streptavidin-binding aptamer sequence 5'-ACCGACCAGAAUCAUGCAAGUGCGUAAGAUAGUCGCGGGCCGGG-3' (*Alu*-S1) was prepared by in vitro transcription using AmpliScribe T7 Flash Transcription kit (Epicentre). To examine self-priming, RNA-free cytoplasmic DNA was prepared from the *Alu*-S1 transfected mouse embryonic carcinoma cells (F9) as described elsewhere in the methods. The presence and identity of reverse transcribed *Alu*-S1 was determined by sequencing of the PCR product using an Alu specific forward primer and an S1 aptamer specific reverse primer.

Alu retrotransposition reporter assay

Retrotransposition reporter assays were carried out as follows. Briefly, 2×10^5 HeLa-HA cells were plated in 6-well tissue culture dish and, one day later, were transfected in triplicate using FuGene 6 (Promega) with 1 µg of the wild type L1 reporter plasmid pJM101/L1.3Δneo as described previously (22), along with 1 µg of *Alu* retrotransposition indicator construct *Alu* neo (gift of J.V. Moran, University of Michigan Medical School) and *Alu* RNA with G25C/G159C mutations. After 72 h, cells were provided DMEM containing 600 µg/mL G418 and 100 µg/mL penicillin/streptomycin (Cellgro). Fourteen days later, the plates were washed with methanol, Giemsa stained, and photographed. Colonies were counted manually using ImageJ (NIH). A total of 1 µg of *Alu*-neo with 1 µg empty driver vector was used as a negative control.

RNA-DNA hybrid assays

To monitor the association of *Alu* RNA and *Alu* cDNA, MEF cells transfected with biotinylated *Alu* RNA (vs. mock) with or without 3TC treatment were used. Briefly, biotinylated *Alu* RNA-transfected cells were UV crosslinked and lysed to collect cytosolic fractions. Streptavidin Dynabeads blocked with 1% BSA (for 18 h) were incubated with the cytosolic lysates diluted in binding buffer (0.2 M Tris-HCl, pH 7.4, 0.5 M NaCl, 20 mM $MgCl_2$, 1% Tween 20). Incubates were gently rotated at 4°C overnight. Then the unbound fraction was collected and the bound fraction was eluted from magnetic beads with elution buffer (pH 2.8) at 70°C for 10 min. All unbound and bound fractions were directly subjected to real-time PCR based *Alu*-cPCR without reverse transcription.

Immunofluorescence staining

For confirmation of in vivo enforced expression of pORFeus-Hs or rat pORF2, mouse RPE/choroid flat mounts were fixed with 2% and stained with FITC-conjugated anti-FLAG F4049 (10 µg/ml, Sigma-Aldrich) or Dylight™ 549 Conjugated anti-V5 antibody (1:5000, Rockland, Cat# 600-442-378). To monitor RNA-DNA hybrids, primary human RPE cells were seeded on chamber slides (Thermo Fisher Scientific, Cat#154941) and transfected with 12.5 pmol *Alu* RNA for 2 h, and then washed with fresh medium to be collected with 2% PFA fixation at the indicated time points. Cells were blocked and incubated with S9.6 antibody (1:200, Kerafast, Cat# ENH001), followed by incubation with secondary antibody and DAPI before image acquisition using confocal microscopy (A1R Nikon).

Pull-down assays

To monitor the association of L1 ORF2p with *Alu* RNA and *Alu* cDNA, RNaseH-deficient HeLa cells (gift of A.P. Jackson and M.A. Reijns) (23) transfected with V5-tagged rat L1 ORF2p plasmid and biotinylated *Alu* RNA (vs. mock) were utilized. Briefly, biotinylated *Alu* RNA-transfected cells were crosslinked with 1% formaldehyde for 15 min at room temperature and lysed to collect cytosolic fractions. Streptavidin Dynabeads blocked with 1% BSA (for 18h) were incubated with the cytosolic lysates diluted in BC200 buffer (20 mM HEPES, pH 7.9, 0.2 mM EDTA, 0.5 mM DTT, 20% glycerol, 0.2% NP-40, and 200 mM KCl) for 2 h at 4°C. The beads were then washed twice with BC200, heated at 95 °C for 30 min. The pull-down samples were subsequently analyzed by immunoblotting with anti-V5 antibody.

Using the same experimental system, a reverse pulldown assay was performed to detect the presence of *Alu* RNA and *Alu* cDNA in the V5-L1 ORF2p immunoprecipitate. Briefly, cytosolic lysates (1 mg at 500 ng/µl) in BC200 buffer, prepared as above, were precleared by incubating with beads for 6 h. Precleared cytosolic lysates were subjected to immunoprecipitation using 15 µg of anti-V5 antibody for overnight at 4 °C. The immune complexes were captured by incubation with pre-blocked beads (4 h at 4 °C). The beads-captured immune complexes were washed twice with BC200. Finally, the beads-captured immune complexes were resuspended in either 1 ml of Trizol reagent followed by RNA purification (for detecting *Alu* RNA) or subjected to Proteinase K treatment and reverse crosslinking (overnight) followed by ethanol precipitation of DNA (for detecting *Alu* cDNA). The *Alu* RNA was detected by direct blotting of biotinylated *Alu* RNA. DNA purified from these assays was analyzed by equator blotting to detect *Alu* cDNA.

Mouse platelet isolation

Platelets used for in situ hybridization were freshly isolated from WT mice as previously described (24). Platelets were resuspended at 1×10^8/mL in serum-free M199 medium and placed on fibrinogen-coated chamber slides (Nunc™ Lab-Tek™ II CC2™ Chamber Slide System; ThermoScientific). Absence of contamination by leukocytes was confirmed using antibodies against CD41 (1:50, Bioss Antibodies, Cat# bs-2636R-A555) and DAPI nuclear staining.

Sequencing

RPE cells were lysed with gentle extraction buffer prepared in 1× PBS containing 1% v/v Triton X-100 (Sigma-Aldrich) and 1 mM EDTA for 15 min on ice. Lysate was centrifuged at 1000g for 10 min at 4°C to pellet-out nuclei. The lysate supernatant was used as the cytoplasmic fraction. Cytoplasmic samples were size-fractionated on a Blue Pippin device (Sage Science) to exclude large molecular weight DNA > 1500-nt. Blue Pippin samples were enriched for ssDNA as determined by Qubit for ssDNA and dsDNA pre- and post-fractionation. Pippin-fractionated ssDNA samples were converted to dsDNA by the Seq Plex Enhanced DNA Amplification Kit (Sigma-Aldrich, SEQXE) without additional fragmentation to enrich for DNA fragments ranging from 200-800 nts. Amplification was monitored by RT-PCR, and cycle number (20–25 cycles) was set as 2-3 cycles after the amplification plateau, as suggested by the manufacturer. 1 µg of dsDNA library was prepared for sequencing with the NEXTflex® Rapid DNA Sequencing Kit (Bioo Scientific). Samples were sequenced on the HiSeq 2500 SE50 platform. The quality of reads was assessed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The reads were then mapped to the human reference genome (hg19: chr1-22, X, Y, M) by two methods: MapSplice (25) and STAR (26). Both of these methods were configured to report the best alignment of each read with minimal mismatches. Only uniquely aligned reads were retained for further analysis. The GTF file containing the genomic locations of all *Alu* species and their family classifications was obtained from UCSC Genome Browser (https://genome.ucsc.edu/cgi-bin/hgTables). Taking the read alignment and the GTF file as input, FeatureCounts (27) was used to calculate the total read count obtained for each *Alu* subfamily.

Read alignment and *Alu* expression analysis

Custom sequence analysis code was written in Python 3.6 and R 3.4.4. Read alignment and feature mapping were executed with STAR 2.5.3a and featureCounts 1.6.1, respectively. In the first step of the pipeline, reads from raw FASTQ files were aligned to the annotated human genome hg19 using alignment software STAR with the following command line call:
./STAR –runThreadN 16 –genomeDir /*path/to/STAR/genome* –outFilterMultimapNmax 40 --readFilesIn /*path/to/FASTQ/file.fq*
The *outFilterMultimapNmax* parameter in the command line above allows us to keep multi-aligned reads with up to 40 alignments. Taking this file, as well as a GTF file storing all annotated, human *Alu* loci, as input, the program featureCounts was adopted to identify reads that are uniquely mapped to each *Alu* locus:
./FC -R SAM -T 16 -t exon -g gene_id -a */path/to/GTF/alu.gtf* -o /*path/to/output/Alu.counts /path/to/alignments/STARAlignments.sam*

Beyond uniquely mapped reads (those are mapped to only one locus on the genome, indicated by the tag 'NH:i:1' in the STAR alignment SAM file), we performed additional analysis of the multi-aligned reads that were family-specific in the *Alu* expression quantification. This was done by a script built in-house. For each read that was multi-mapped (mapped equally well to more than one *Alu* insertion), the script iterated through all of its mapped *Alu* locations and determined whether it was mapped to a single family or multiple families. Our data contained about 325k uniquely aligned *Alu* reads and about 25k multi-aligned reads that were family-specific. Over 98% of these reads are mapped with no more than 2 mismatches to the reference sequences. The overall distribution of family-level *Alu* abundances taking into account family-specific multi-mapped reads in addition to uniquely mapped reads were plotted. To explore whether reads could be uniquely mapped to the young *Alu*Y family, we collected reads that mapped to individual *Alu*Y subfamilies and also identified reads that were multi-mapped to both the *Alu*Y family and at least one of its young *Alu* subfamilies. In total, there were 2,549 family-specific young *Alu*Y reads and 1,655 reads that could be mapped to both *Alu*Y and its young *Alu* families.

Code Availability

Example scripts illustrating the sequence analysis process are publicly hosted at https://github.com/ElkHairCaddis/Alu.

Statistical analysis

The binary readouts of RPE degeneration (i.e., the presence or absence of RPE degeneration on fundus and ZO-1-stained flat-mount images) were analyzed with Fisher's exact test. Cell-morphometry data were assessed with two-tailed Mann-Whitney U test. All other data were expressed as means ± s.e.m. and were analyzed with Mann-Whitney U test. P values <0.05 were deemed statistically significant. Sample sizes were selected on the basis of power analysis $\alpha = 5\%$; $1 - \beta = 80\%$, such that we were able to detect a minimum of 50% change, assuming a sample s.d. based on Bayesian inference. Outliers were assessed with Grubbs' test. On the basis of this analysis, no outliers were detected, and no data were excluded. Fewer than 5% of subretinal-injection recipient tissues were excluded on the basis of predetermined exclusion criteria (including hemorrhage and animal death due to anesthesia complications) relating to the technical challenges of this delicate procedure.

Health Records Database Analyses

Participants and Sample Selection

Subjects were included in the analysis if they were 50 years or older and had 1 or more record of a prescription medication. Subjects were assigned to the NRTI user group if they had 1 or more record of NRTI prescription for preexposure prophylaxis (PrEP) and to the NRTI non-user group if they had no records of NRTI prescription. The index date for the PrEP cohort was the date of first NRTI prescription, while for controls (NRTI non-users) the index date was the first date of any medication. Subjects were excluded if they had a diagnosis of HIV or hepatitis B or if they were younger than 50 years of age. For the main analysis, individuals with pre-existing atrophic ("dry") age related macular degeneration (≥1 medical claim prior to index date) were excluded. For Falsification Test 1, individuals with pre-existing appendicitis (≥1 medical claim prior to diagnosis of HIV or hepatitis B) were excluded. For Falsification Test 2, individuals with pre-existing hernia (≥1 medical claim prior to diagnosis of HIV or hepatitis B) were excluded.

Exposure to NRTI

Individuals were classified as receiving NRTI if they filled ≥1 outpatient pharmacy prescription for an NRTI as extracted from the outpatient VA pharmacy data or US National Drug Codes. Exposure to NRTI medication was the key predictor and was set as 1 if a patient had any exposure to an NRTI during the study and 0 otherwise.

Dependent Variable

Time to initial diagnosis of dry AMD, appendicitis, or hernia during follow-up period, as identified by the ICD-9-CM codes corresponding to dry AMD (362.51), appendicitis (540-542), and hernia

(550-553), and ICD-10-CM codes corresponding to dry AMD (H35.31), appendicitis (K35-K37), and hernia (K40-K44), was the dependent variable for the analyses. Observation of beneficiaries was right censored at study end.

Analyses

Analyses were performed with the use of SAS software, version 9.4 (SAS Institute) and R software, version 3.6.1 (the R project [https://www.r-project.org]). To analyze the risk of dry AMD, appendicitis, or hernia between those exposed to NRTIs and those not exposed to NRTI medications, we fit Cox proportional hazard models adjusted for these covariates: age, gender, race (available in Veterans and Humana), body-mass index (BMI), tobacco use, and Charlson comorbidity index score. For the PearlDiver Humana and PearlDiver Mariner databases, we used classic Cox proportional hazard models. The restricted maximum-likelihood estimator method was used to estimate the between-study variance.

Because there were a low number of incident dry AMD cases (28) among NRTI users in the Truven and Veterans databases, we employed Bayesian proportional hazards Cox regression models for those database analyses using the R package spBayesSurv. The Bayesian model is specified with the prior distribution:

$$\beta \sim N_p(\beta_0, S_0),$$
$$h_k | h \overset{iid}{\sim} \Gamma(r_0 h, r_0),\ k = 1,...,M,$$
$$(\theta_1, \theta_2) \sim \text{Beta}(\theta_{1a}, \theta_{1b}) \times \Gamma(\theta_{2a}, \theta_{2b}).$$

with hyperparameters: M = 10, $r_0 = 1$, $h = \hat{h}$, $\beta_0 = 0$, $S_0 = 10^5 I_p$, $\theta_0 = (\theta_{1a}, \theta_{1b}, \theta_{2a}, \theta_{2b})^T = (1,1,1,1)^T$, where M is the number of pieces of the hazard function, r_0 is the weight given to the exponential distribution underlying the baseline hazard function, and $\hat{h}$ is the estimated rate parameter from an exponential proportional hazards model fit via maximum likelihood (29). We also performed sensitivity analysis for the prior specification changing the hyperparameter to M = 20 and using r_0 values of 0.1, 1, or 100. When r_0=0.1, very little weight is given to the centering exponential distribution, whereas r_0=1000 essentially fits an exponential regression model. The intermediate values of r_0=1 allows extrapolation into the tails.

Because of computational intensity, for the Truven, Veterans, and PearlDiver Mariner datasets, the study population analyzed for the Cox regression models comprised all the NRTI users and a random sample of the NRTI non-users. We present estimated mean, and 95% lower and upper values of the adjusted hazard ratios. Statistical tests were two-sided. *P* values of less than 0.05 were considered to indicate statistical significance.

Propensity score matching

To further evaluate the robustness of our findings and to mitigate any possible residual confounding, we estimated propensity-score models including use of NRTIs and no use of NRTIs. The individual propensities for starting NRTI treatment were estimated with the use of logistic regression. As predictors, the propensity-score models included the set of variables which displayed P values < 0.1 in logistic regression analyses. In the Veterans Database, these variables were thus used in the propensity score model: age, gender, race, body mass index, tobacco use, and Charlson comorbidity index. In the Truven Marketscan Database, these variables were thus used in the propensity score model: age, gender, Charlson comorbidity index, body mass index, and smoking. In the PearlDiver Humana Database, these variables were thus used in the propensity score model: age, gender, body mass index, and Charlson comorbidity index. In the PearlDiver Mariner Database, these variables were thus used in the propensity score model: age, gender, body mass index, and Charlson comorbidity index. We used the R package MatchIt to perform matching in a 1:1 ratio using greedy nearest neighbor matching. In addition, to control for any residual covariate imbalance, we estimated the relative hazard in the propensity score-matched groups using the multivariable classic or Bayesian Cox models that included the covariates from the multivariable regression analysis employed for the original

unmatched group analyses. Statistical tests were two-sided. P values < 0.05 were considered statistically significant.

Meta-analysis (Frequentist)
For the primary analysis, an inverse-variance weighted meta-analysis of the four databases was performed to estimate the combined hazard ratio (HR) and to compute 95% confidence intervals (CI) using both fixed-effect and random-effects models. Meta-analyses were performed using the R package metafor. The restricted maximum-likelihood (REML) estimator was used to estimate the between-study variance (t^2). Variability among the five databases was evaluated using Cochran's Q-test (30). A random-effects model was used in the primary analyses as it assumes that individual databases are samples of different populations with different underlying true effects. In contrast, fixed-effect models assume that individual databases are samples from the same population (31–33). A forest plot was created to depict the HR and 95% CI of each study and of the pooled results.

Meta-analysis (Bayesian)
As a secondary analysis, Bayesian meta-analysis was performed using a random-effects normal-normal hierarchical model (the same as the random-effects model above), which accounts for uncertainty in estimating the between-study variance. For the effect parameter μ, we choose a neutral unit information prior given by a normal prior with mean $\mu_p = 0$ (centered around a hazard ratio of 1.0) and a variance of ($\sigma_p^2 = 4$) (34). In a hierarchical model $\theta \sim N[\mu, \tau^2]$, where τ^2 is the between study variance of the logarithmic HR, the "range" of HRs, defined as the ratio of the 97.5% and 2.5% quantiles of the HR, is equal to ($e^{3.92\tau}$) (34). We used a weakly informative half-Cauchy prior distribution (35, 36) for between-study variability with the assumption that it was unlikely for the between-study HRs to vary by more than 3-fold. For this assumption, range = 3 and $\tau = 0.280$ (scale). We performed a sensitivity analysis to the choice of the prior by assuming that it was unlikely for the between-study HRs to vary by more than 10-fold. For this assumption, range = 10 and $\tau = 0.587$ (scale). Meta-analyses were performed using the R package bayesmeta. A forest plot was created to depict the HR, the shrinkage estimates, and 95% credible interval of each study and of the summary results. Computation of posterior predictive P values were implemented in bayesmeta via Monte Carlo sampling.

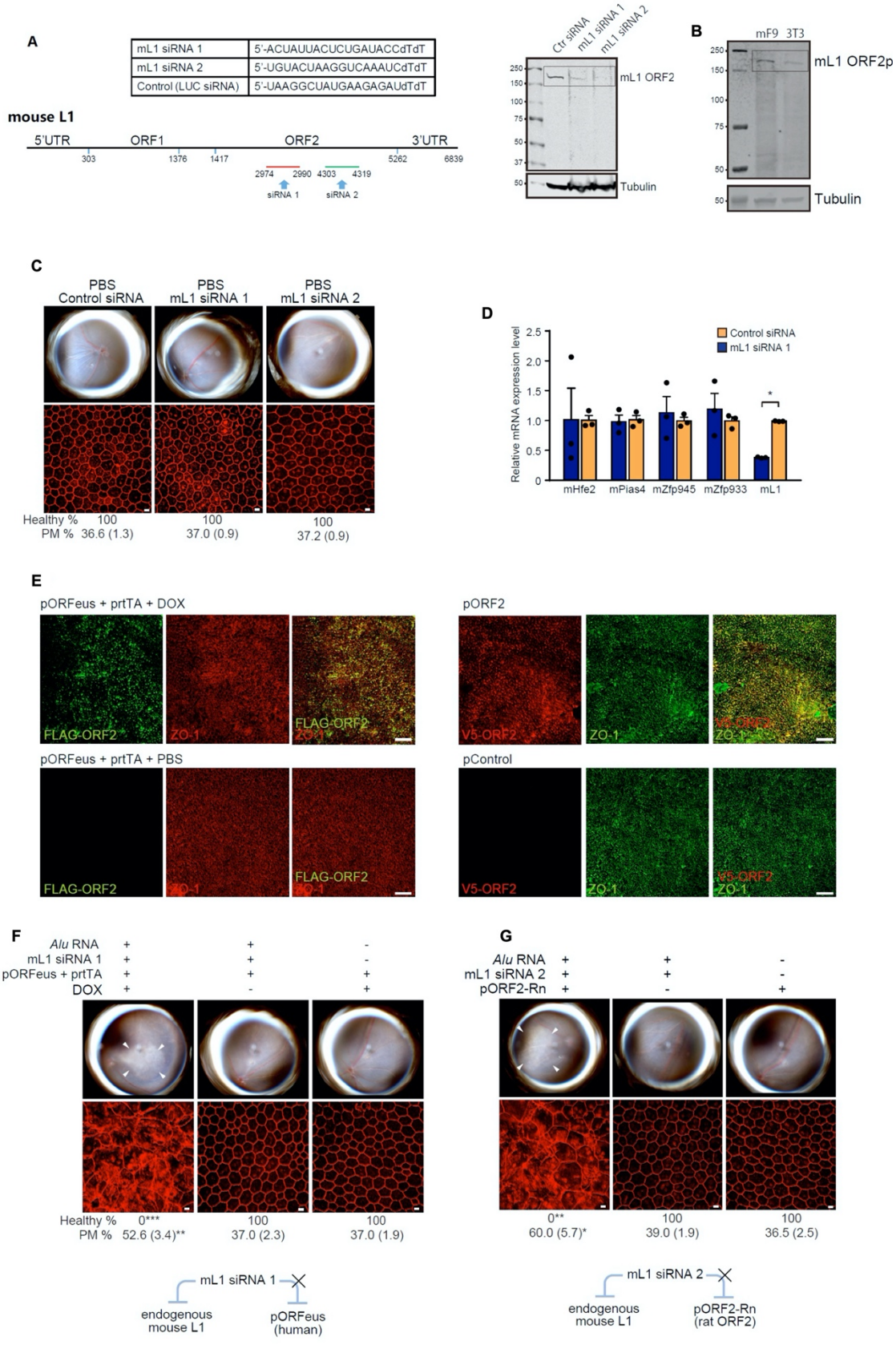

A
mL1 siRNA 1
5'-ACUAUUACUCUGAUACCdTdT
mL1 siRNA 2
5'-UGUACUAAGGUCAAAUCdTdT
Control (LUC siRNA)
5'-UAAGGCUAUGAAGAGAUdTdT
mouse L1
5'UTR
ORF1
ORF2
3'UTR
303
1376
1417
2974
2990
4303
4319
5262
6839
siRNA 1
siRNA 2
Ctr siRNA
mL1 siRNA 1
mL1 siRNA 2
mL1 ORF2
Tubulin
B
mF9
3T3
mL1 ORF2p
Tubulin
C
PBS
Control siRNA
PBS
mL1 siRNA 1
PBS
mL1 siRNA 2
Healthy %
PM %
100
36.6 (1.3)
100
37.0 (0.9)
100
37.2 (0.9)
D
Control siRNA
mL1 siRNA 1
Relative mRNA expression level
mHfe2
mPias4
mZfp945
mZfp933
mL1
E
pORFeus + prtTA + DOX
pORF2
FLAG-ORF2
ZO-1
V5-ORF2
pORFeus + prtTA + PBS
pControl
F
Alu RNA
mL1 siRNA 1
pORFeus + prtTA
DOX
Healthy %
PM %
0***
52.6 (3.4)**
100
37.0 (2.3)
100
37.0 (1.9)
mL1 siRNA 1
endogenous mouse L1
pORFeus (human)
G
Alu RNA
mL1 siRNA 2
pORF2-Rn
0**
60.0 (5.7)*
100
39.0 (1.9)
100
36.5 (2.5)
mL1 siRNA 2
endogenous mouse L1
pORF2-Rn (rat ORF2)

Fig. S1. mL1 targeting. (A) Sequence and location of siRNAs specific for mL1 (left) and immunoblots of mouse RPE cells (right) using an anti-mouse L1 ORF2p antibody. **(B)** Immunoblotting analysis of endogenous mouse L1 ORF2p from mouse F9 embryonal carcinoma cells and NIH-3T3 cells, which contain high and low abundance of L1 ORF2p, respectively. Tubulin expression is presented as loading control. **(C)** Fundus photos (top) and zonula occludens-1 (ZO-1, red)-stained RPE flat mounts(bottom) of wild-type (WT) mice injected with PBS, control (Luc) siRNA, or mL1 siRNA. Regular hexagonal cellular boundaries represents normal RPE. Scale bars, 10 μm. n = 6–11. Binary and morphometric quantification of RPE are shown PM, polymegethism (mean (SEM)). **(D)** Quantification of gene expression by real-time PCR, normalized to 18S rRNA abundance, in WT mouse RPE cells transfected with mouse L1 (mL1) siRNA or control (Luc) siRNA. Genes with the greatest BLAST sequence matches to the mL1 siRNA were also quantified. n = 3. *P < 0.05 by Mann-Whitney U test. Error bars show SEM. **(E)** After subretinal transfection of FLAG-tagged pORFeus-Hs (encoding a synthetic, codon optimized full-length human L1 element) along with prtTA (encoding reverse tetracycline-controlled transactivator) and doxycycline (DOX) or PBS, or after subretinal administration of V5-tagged rat pORF2 in WT mice, RPE flat mounts were stained with anti-FLAG antibody (left) or anti-V5 antibody (right) along with anti ZO-1 antibody. Scale bars, 100 μm. **(F** and **G)** RPE degeneration (fundus photos, top, and ZO-1 flat mounts, bottom) in WT mice after subretinal transfection of mL1 siRNA-refractory L1 constructs. Schematic of experiments shown at bottom. White arrowheads indicate degenerated area in the fundus photos. **(F)** RPE degeneration in the presence or absence of *Alu* RNA, L1 siRNA #1, an siRNA-refractory pORFeus-Hs encoding a synthetic full-length human L1 that is activated by administration of prtTA (reverse tetracycline-controlled transactivator) and doxycycline (DOX). **(G)** RPE degeneration in the presence or absence of *Alu* RNA, L1 siRNA #2, and pORF2-Rn encoding siRNA-refractory rat L1 ORF2p. *P < 0.05; **P < 0.01; ***P < 0.001, Fisher's exact test for binary; two-tailed t-test for morphometry.

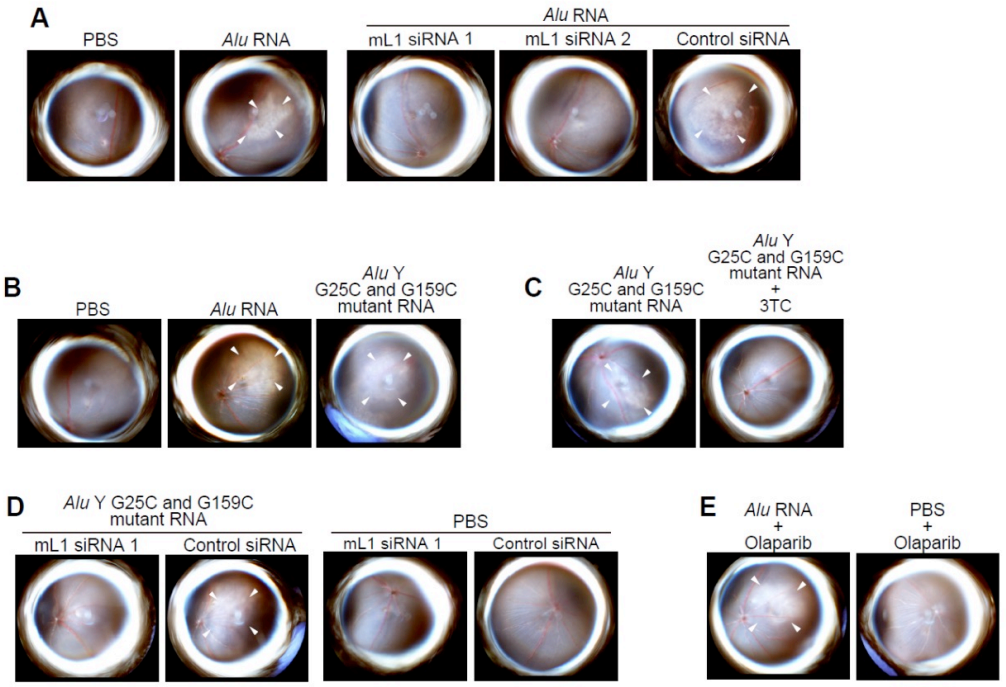


Fig. S2. Endogenous L1 is required for *Alu* RNA-induced RPE toxicity. (A) Fundus photographs representing the respective RPE sheets of mice from Figure 1. Arrowheads in fundus images denote the boundaries of RPE hypopigmentation. Fundus photographs in wild-type (WT) mice administered *Alu* RNA or PBS, and *Alu* RNA with either of two L1-targeted siRNAs or control siRNA. *n* = 6–15. **(B)** Fundus photographs in WT mice following administration of *Alu* RNA or *Alu* G25C/G159C double mutant RNA. *n* = 6. **(C)** *Alu* G25C/G159C double mutant RNA-induced RPE degeneration in WT mice was blocked by 3TC. *n* = 5–6. **(D)** *Alu* G25C/G159C double mutant RNA or PBS subretinal injection into WT mice with mL1 siRNA or control siRNA. *n* = 6–11. **(E)** Fundus photographs in WT mice treated with *Alu* RNA and olaparib, a chemical inhibitor of L1 retrotransposition. *n* = 4–7.

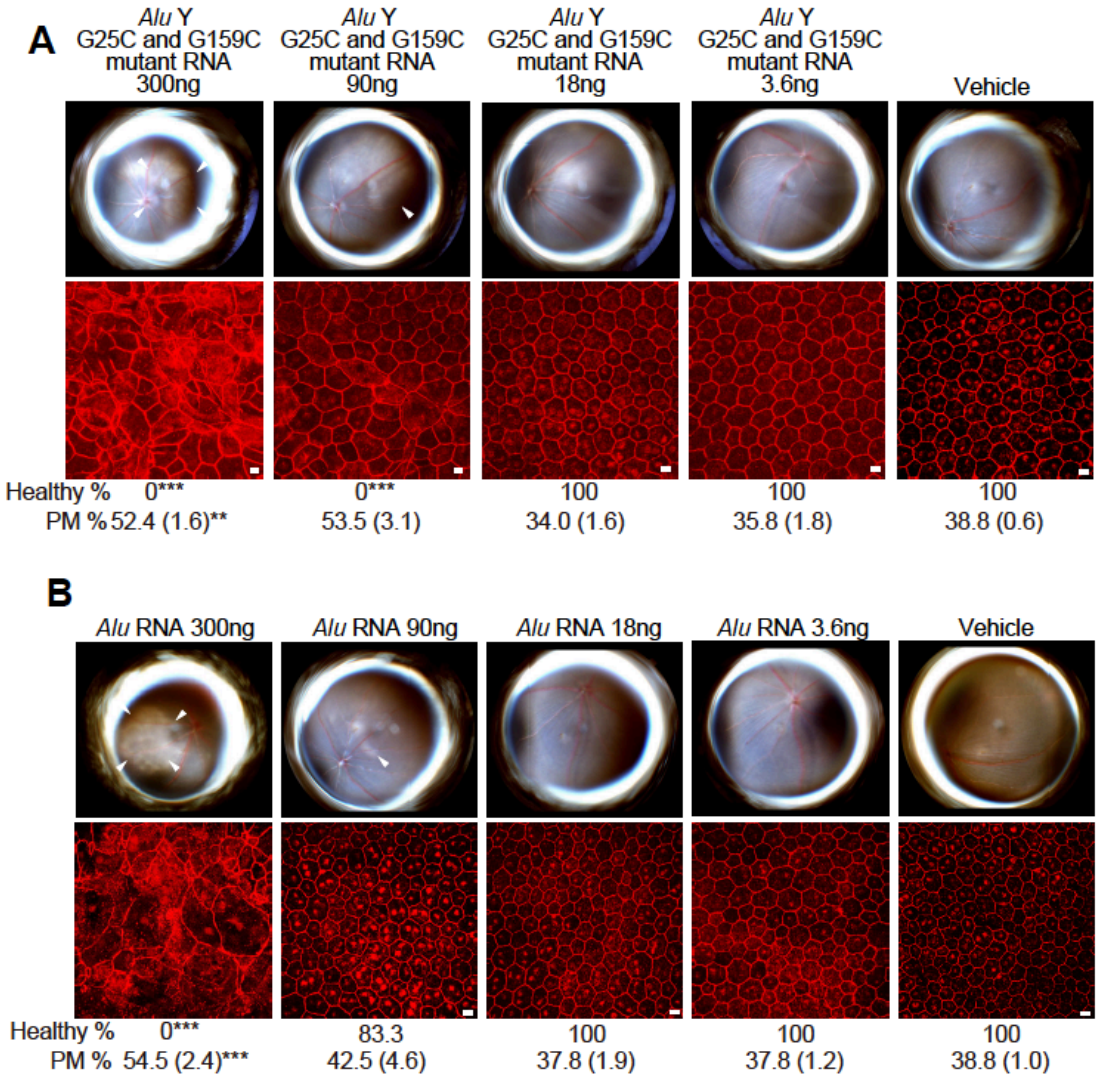


Fig. S3. Toxicity of *Alu* containing G25C/G159C mutations (*Alu* G25C/G159C double mutant RNA). (A and **B)** RPE degeneration (fundus photographs, top; ZO-1 flat mount micrographs, bottom) in WT mice. Binary and morphometric quantification of RPE degeneration are shown (***P* < 0.01; ****P* < 0.001, Fisher's exact test for binary; two-tailed t-test for morphometry). PM, polymegethism (mean (SEM)). Arrowheads in fundus images denote the boundaries of RPE hypopigmentation. Scale bars, 10 μm. **(A)** Dose-ranging studies of subretinal administration of *Alu* G25C/G159C double mutant RNA. *n* = 4–8. **(B)** Dose-ranging studies of subretinal administration of *Alu* RNA in WT mice. *n* = 6–18.

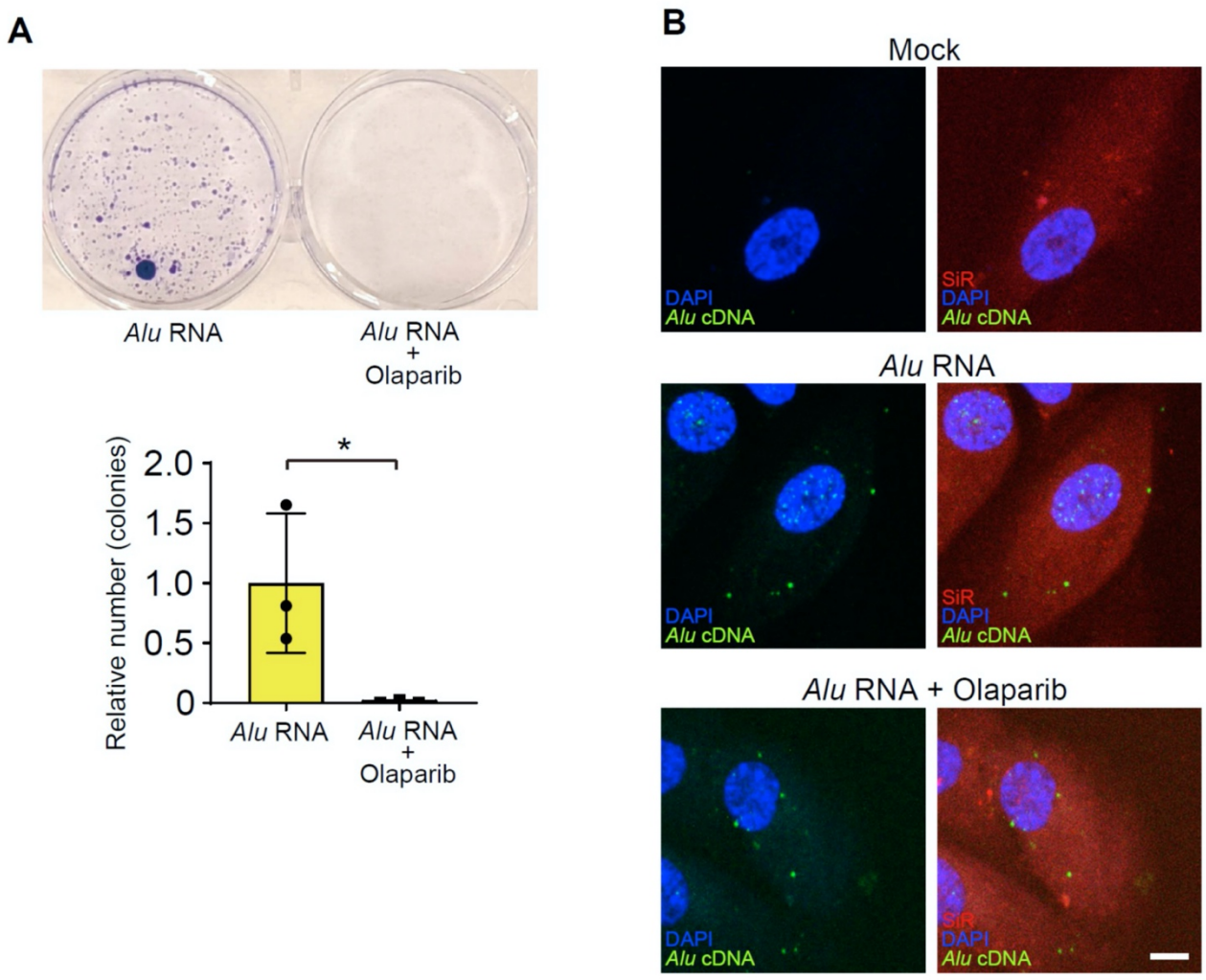


Fig. S4. Olaparib inhibits *Alu* retrotransposition but not reverse transcription of *Alu* RNA into *Alu* cDNA. (A) Retrotransposition of *Alu* RNA in the presence of olaparib in a cellular *Alu* retrotransposition reporter assay (see Supplementary Methods). **P* < 0.05 by Mann-Whitney U test. Error bars show SEM. *n* = 3. **(B)** In situ hybridization of *Alu* cDNA (green) in primary human RPE cells transfected with *Alu* RNA in the presence of olaparib (10 μM). DAPI (blue), SiR (F-actin, red). Scale bars, 10 μm.

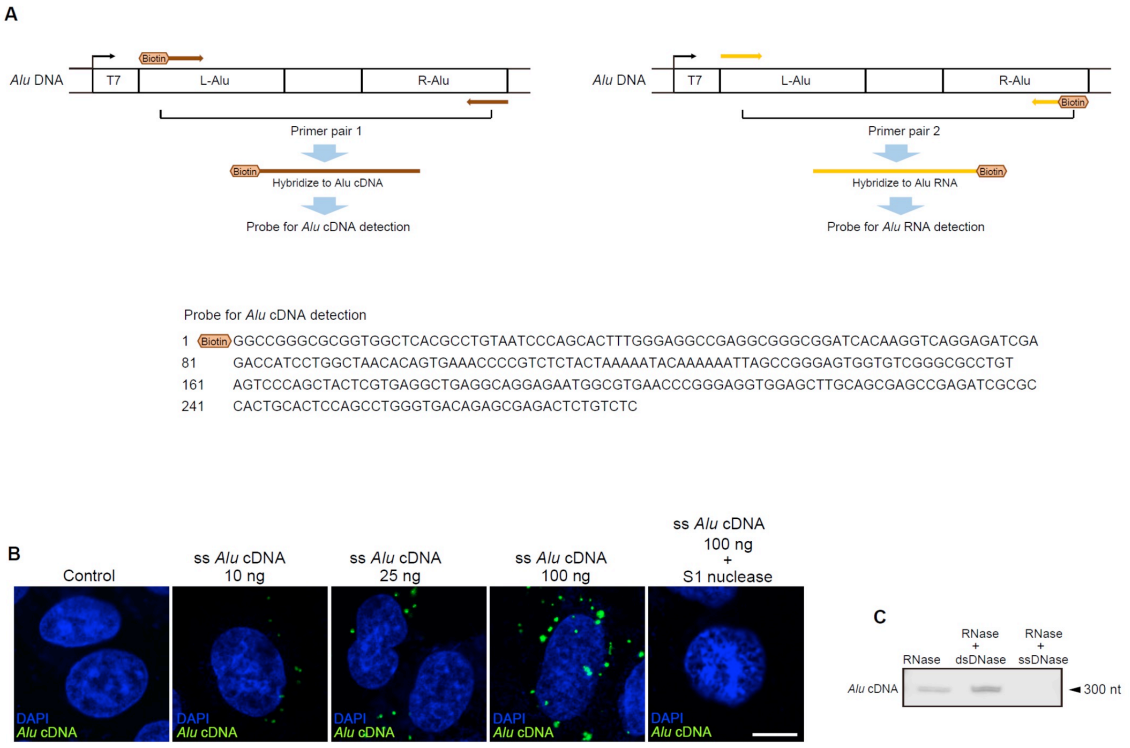


Fig. S5. Specificity of *Alu* cDNA probe. (**A**) Schematic of the generation of the *Alu*-strand specific probe for equator blotting or northern blotting. Primers were selected for PCR to generate two sets of single-stranded, biotin labelled DNA probes; one for *Alu* cDNA and another for *Alu* RNA. Biotinylated single-stranded DNA probe for detection of *Alu* cDNA was synthesized by PCR from a linearized plasmid containing a consensus *Alu* Y element using the *Alu* specific 5'-biotinylated forward and reverse primer. Biotinylated single-stranded DNA probe for detection of *Alu* RNA was synthesized by PCR using *Alu* specific forward and 5'-biotinylated reverse primer and then purified (top). The sequence of the *Alu* cDNA probe is depicted at bottom. **(B)** In situ hybridization of *Alu* cDNA (green) in primary human RPE cells transfected with artificially synthesized single-stranded *Alu* cDNA (ss *Alu* cDNA) with or without S1 nuclease. DAPI (blue), Scale bars, 10 μm. **(C)** *Alu* cDNA blotting in primary human RPE cells. Primary human RPE cells transfected with *Alu* RNA and cytoplasmic nucleic acid and treated with RNase, a double-stranded DNase, or a single-stranded DNase (S1 nuclease).

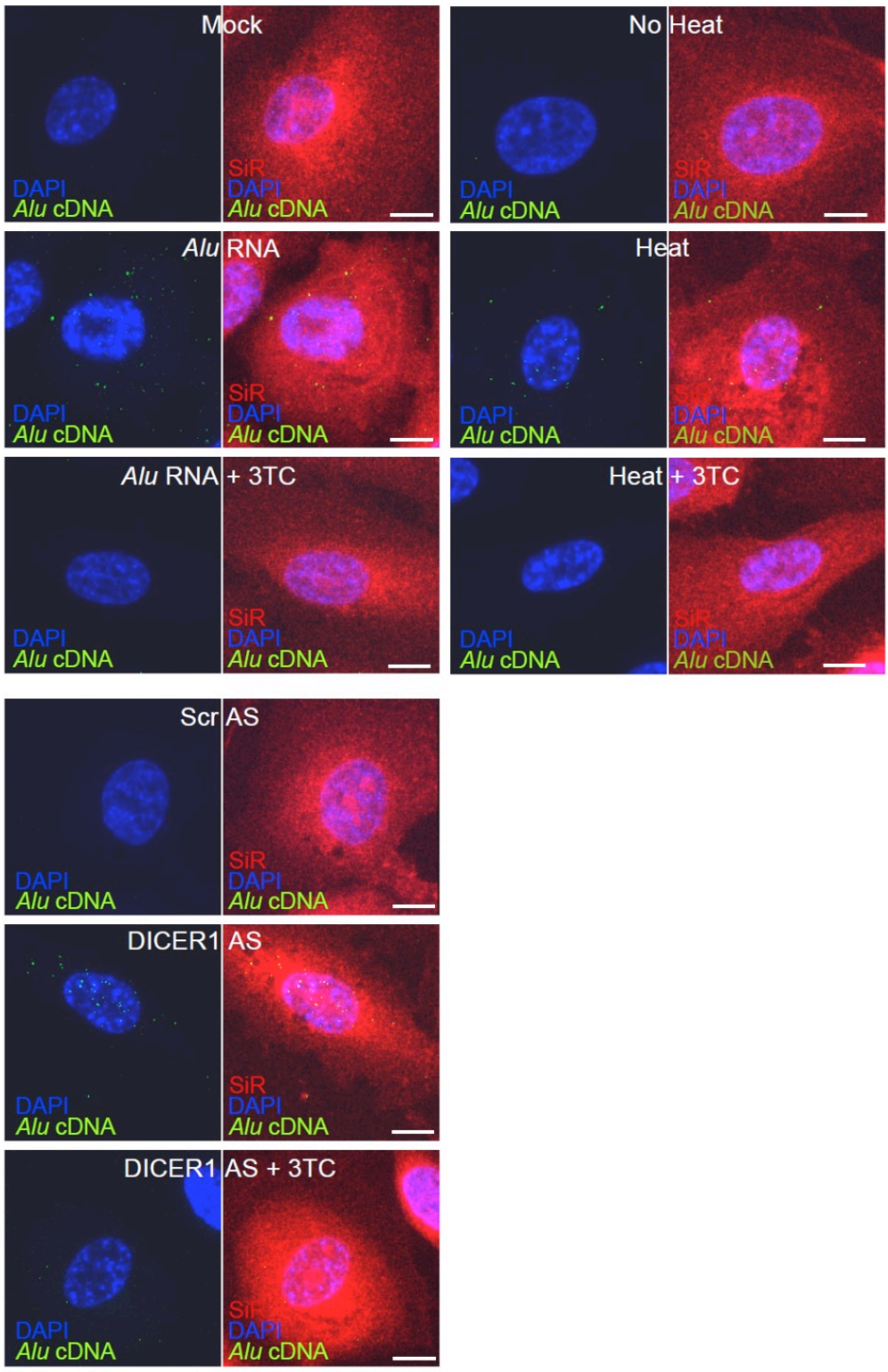


Fig. S6. Endogenous reverse-transcribed cytoplasmic *Alu* cDNA. Fluorescent micrographs of in situ hybridization of *Alu* cDNA in human RPE cells (green) co-labeled with DAPI (blue) to identify nuclei and SiR-actin (red) to identify cytoplasm. Cells were transfected with *Alu* RNA, exposed to heat shock, or transfected with DICER1 antisense oligonucleotides (DICER1 AS) in the presence or absence of 3TC. Representative of n = 3 experiments. Scale bar, 10 μm.

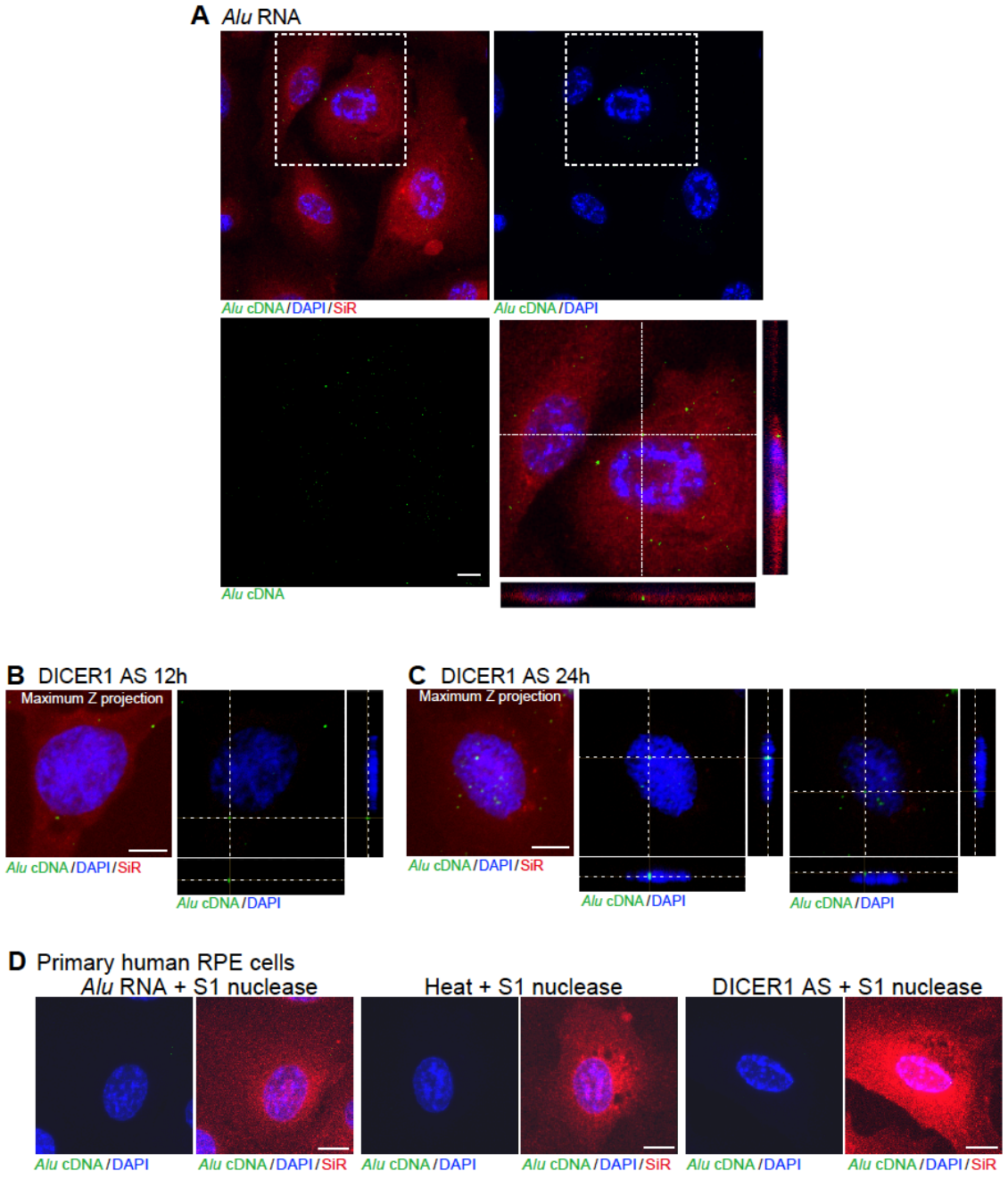


Fig. S7. Endogenous *Alu* cDNA in human RPE cells. (A to **D)** In situ hybridization of *Alu* cDNA (green) in primary human RPE cells. DAPI (blue), SiR (F-actin, red). Scale bars, 10 µm. Orthogonal views obtained by laser scanning confocal microscopy (far-right image) show co-localization of *Alu* cDNA (green) with cytoplasmic F-actin (SiR, red) with DAPI counterstain. **(A)** *Alu* cDNA detection after transfection with *Alu* RNA. **(B** and **C)** In situ hybridization of endogenous *Alu* cDNA after transfection with DICER1 antisense oligonucleotides (DICER1 AS). At 12 h after exposure to DICER1 AS **(B)**, *Alu* cDNA is localized in the cytoplasm. At 24 h **(C)**, *Alu* cDNA accumulation remains predominantly cytoplasmic but was occasionally observed in the nucleus. **(D)** Effect of S1 nuclease on in situ hybridization of *Alu* cDNA in primary human RPE cells after *Alu* RNA transfection (left) or heat shock (middle) or DICER1 AS (right). S1 nuclease was added after cells were fixed, prior to hybridization. Scale bars, 10 µm.

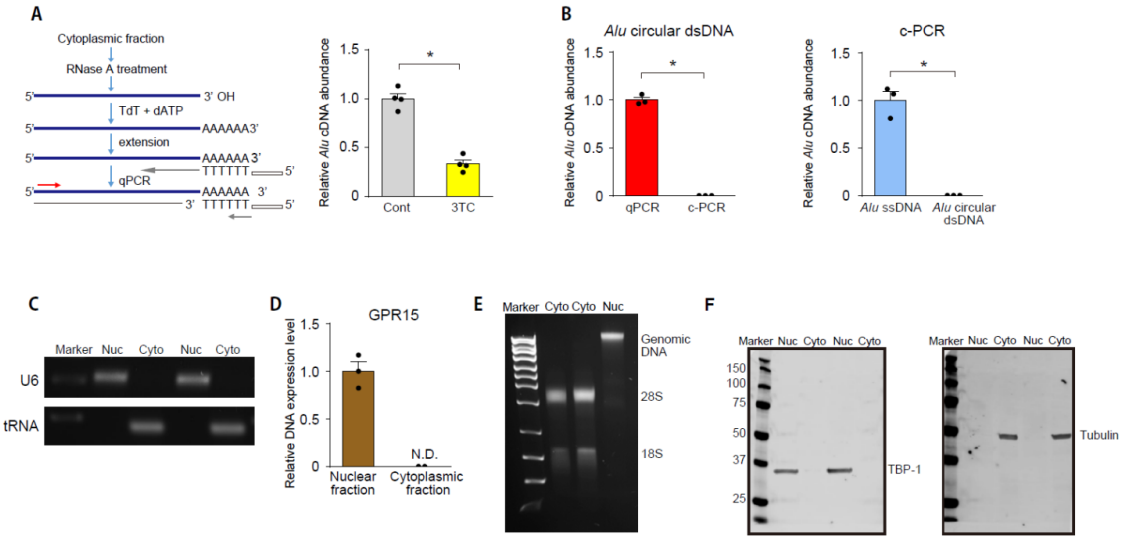


Fig. S8. Reverse transcribed endogenous *Alu* cDNA originating from *Alu* RNA in human cells. (A) Schematic of the method (*Alu* c-PCR) used to purify and amplify reverse transcribed single-stranded DNA (left). Total cell lysate was separated into nuclear and cytoplasmic fractions, and then RNase-treated to improve tailing efficiency. Cytoplasmic DNA was tailed on the 3' end to generate a 20–40 poly A tail by using terminal deoxynucleotidyl transferase (TdT), and then the poly T-anchored primer (TAV oligo) was annealed to the poly A-tail of the template strand and extended. Anchored DNA was amplified using primer specific for the anchor and reverse primer specific for the sequence within *Alu*. Right panel shows endogenous *Alu* cDNA abundance in primary human RPE cells after 3TC treatment. **P* < 0.05 by Mann-Whitney U test. Error bars show SEM. **(B)** *Alu* single-stranded DNA (ssDNA), but not *Alu* circular double-stranded DNA (dsDNA), is amplified by *Alu* c-PCR. **P* < 0.05 by Mann-Whitney U test. **(C)** Real-time RT-PCR for U6 RNA and tRNA to confirm enrichment and lack of cross-contamination in nuclear (Nuc) and cytoplasmic (Cyto) fractions. **P* < 0.05 by Mann-Whitney U test. **(D)** Direct real-time PCR using a primer set for the intron-intron junction of GPR15 showed absence of genomic DNA contamination in the cytoplasmic fraction. Error bars show SEM. N.D., not detected. **(E)** 0.9% agarose gel electrophoresis of cytoplasmic and nuclear RNA isolated from primary human RPE cells shows genomic DNA present in the nuclear fraction but not detected in in two duplicate cytoplasmic fraction samples. **(F)** Enrichment and lack of cross-contamination in nuclear and cytoplasmic fractions by immunoblotting for TBP-1 and tubulin.

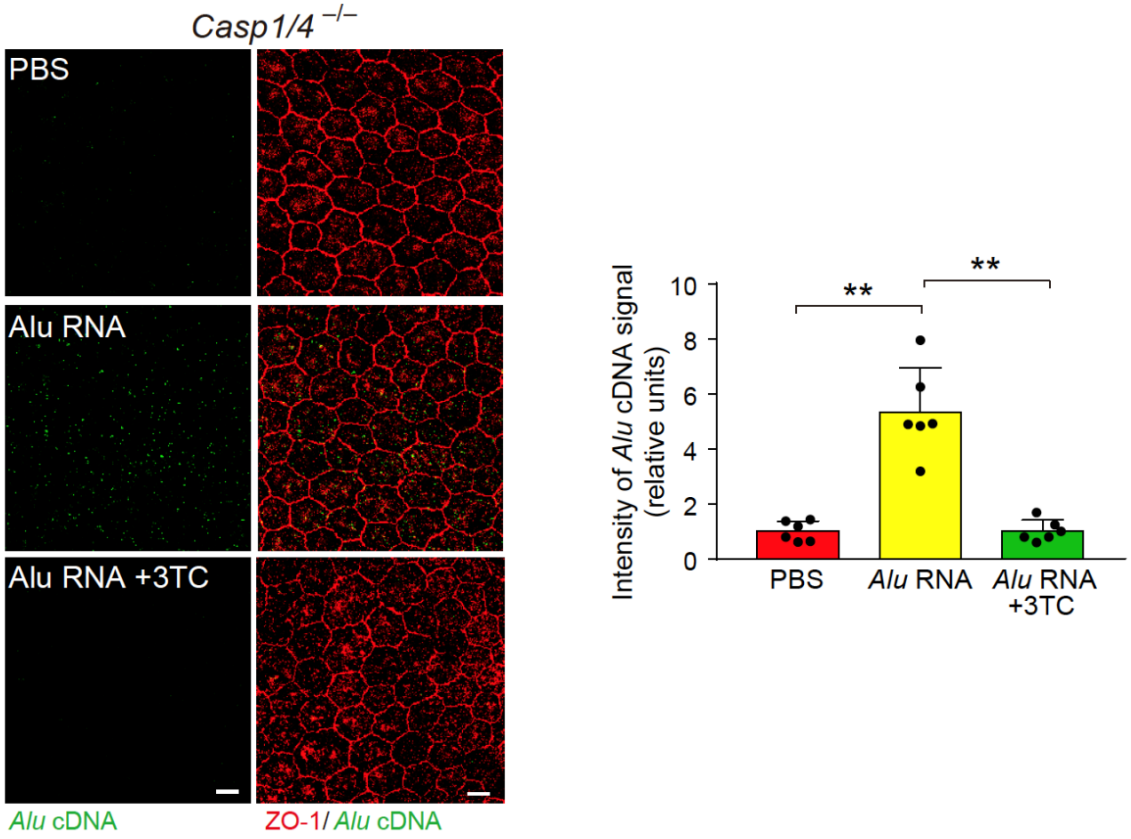


Fig.S9. *Alu* cDNA accumulation in RPE of Casp1/4 dko mice after *Alu* RNA subretinal injection. In situ hybridization to detect *Alu* cDNA (green) in RPE flat mounts 1 day after subretinal *Alu* RNA injection in Casp1/4 dko mice and in wild-type mice treated with 3TC. Zonula occludens-1 (ZO-1, red). Bar graph depicts quantification of *Alu* cDNA signal. $n = 6$. $^*P < 0.01$ by Mann-Whitney U test. Error bars show SEM.

A

hL1 siRNA 1	5'-AUGGAAUAUGAAAUGAAdTdT
hL1 siRNA 2	5'-AAGACACAUGCACACGUAUGUdTdT
Control (Scr siRNA)	5'-UCAAGAAGCCAAGGAUAAdTdT

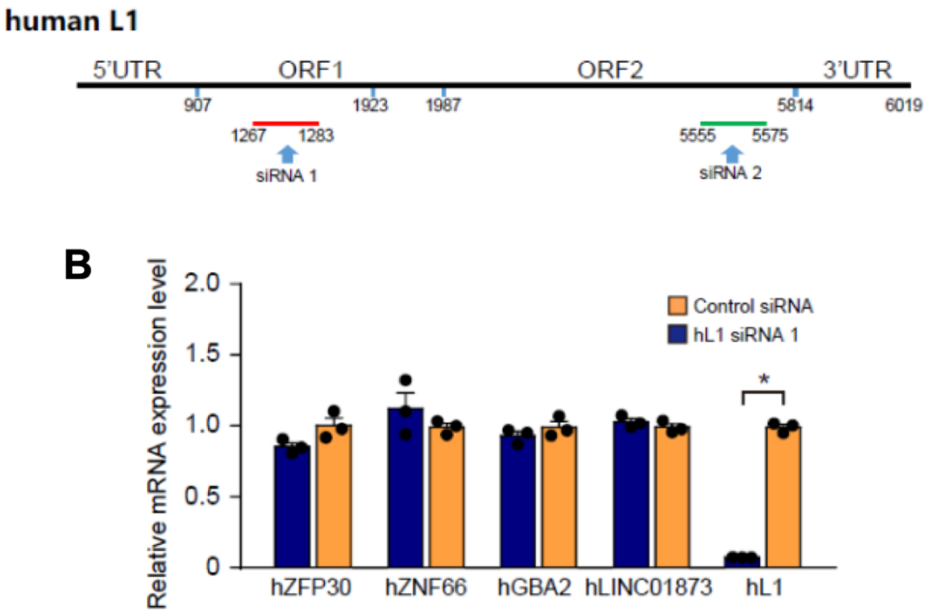


Fig. S10. hL1 targeting. (A) Sequence and location of siRNAs specific for hL1. **(B)** Expression analysis of genes with the greatest BLAST sequence matches to the hL1 siRNA from primary human RPE cells transfected with human L1 (hL1) siRNA or control (Luc) siRNA. Quantification was performed by real-time PCR and normalized to 18S rRNA abundance. $n = 3$. $^*P < 0.05$ by Mann-Whitney U test. Error bars show SEM.

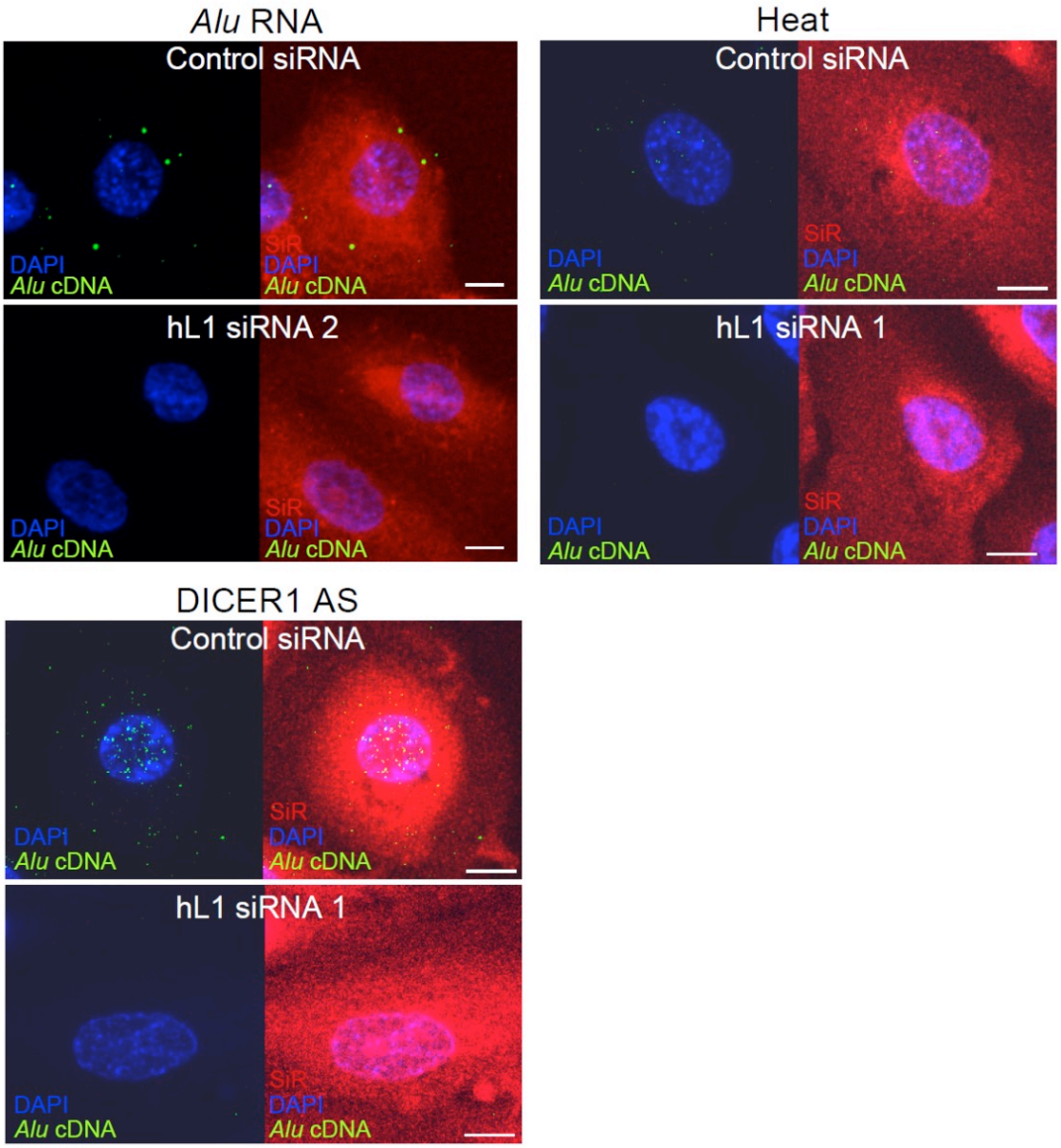


Fig. S11. Endogenous *Alu* cDNA synthesis is L1 dependent. In situ hybridization of *Alu* cDNA (green) in primary human RPE cells after treatment with in vitro transcribed *Alu* RNA, DICER1 antisense oligonucleotides (DICER1 AS), or heat shock after transfection with human (h) L1 siRNA. DAPI (blue), SiR (F-actin, red). Scale bars, 10 μm.

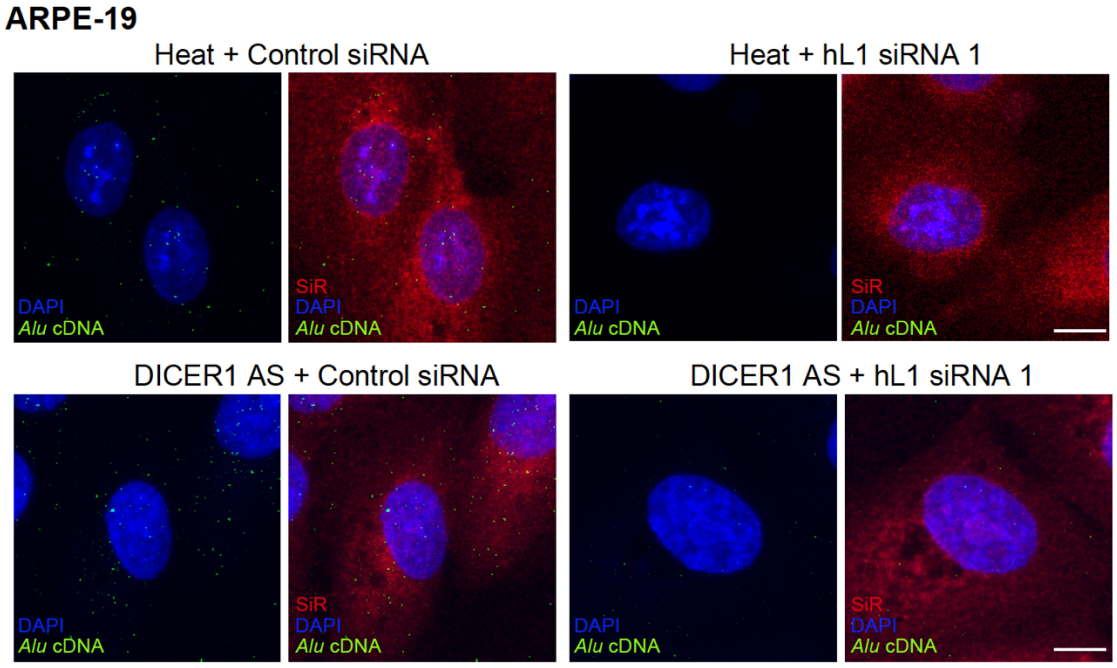


Fig. S12. hL1 siRNA decreases *Alu* cDNA abundance in ARPE-19 cells. In situ hybridization of *Alu* cDNA (green) in ARPE-19 after treatment with heat shock (top) or DICER1 antisense oligonucleotides (DICER1 AS, bottom) after transfection with human (h) L1 siRNA #1 compared to control siRNA. DAPI (blue), SiR (F-actin, red). Scale bars, 10 μm.

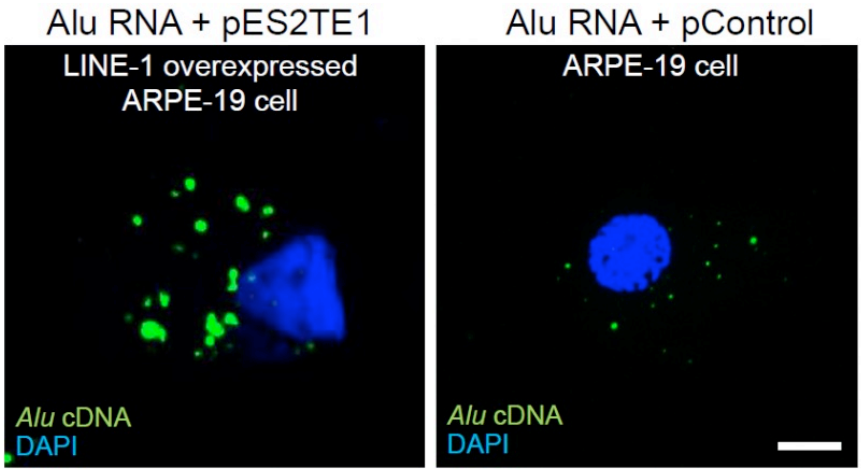


Fig. S13. L1 ORF2p overexpression enhances *Alu* cDNA synthesis following *Alu* RNA transfection. In situ hybridization of *Alu* cDNA (green) in ARPE-19 cells after *Alu* RNA transfection following L1 ORF2p overexpression induced by transfection of pES2TE1 (an engineered human L1 expression vector; 19) or pControl. DAPI (blue). Scale bars, 10 μm.

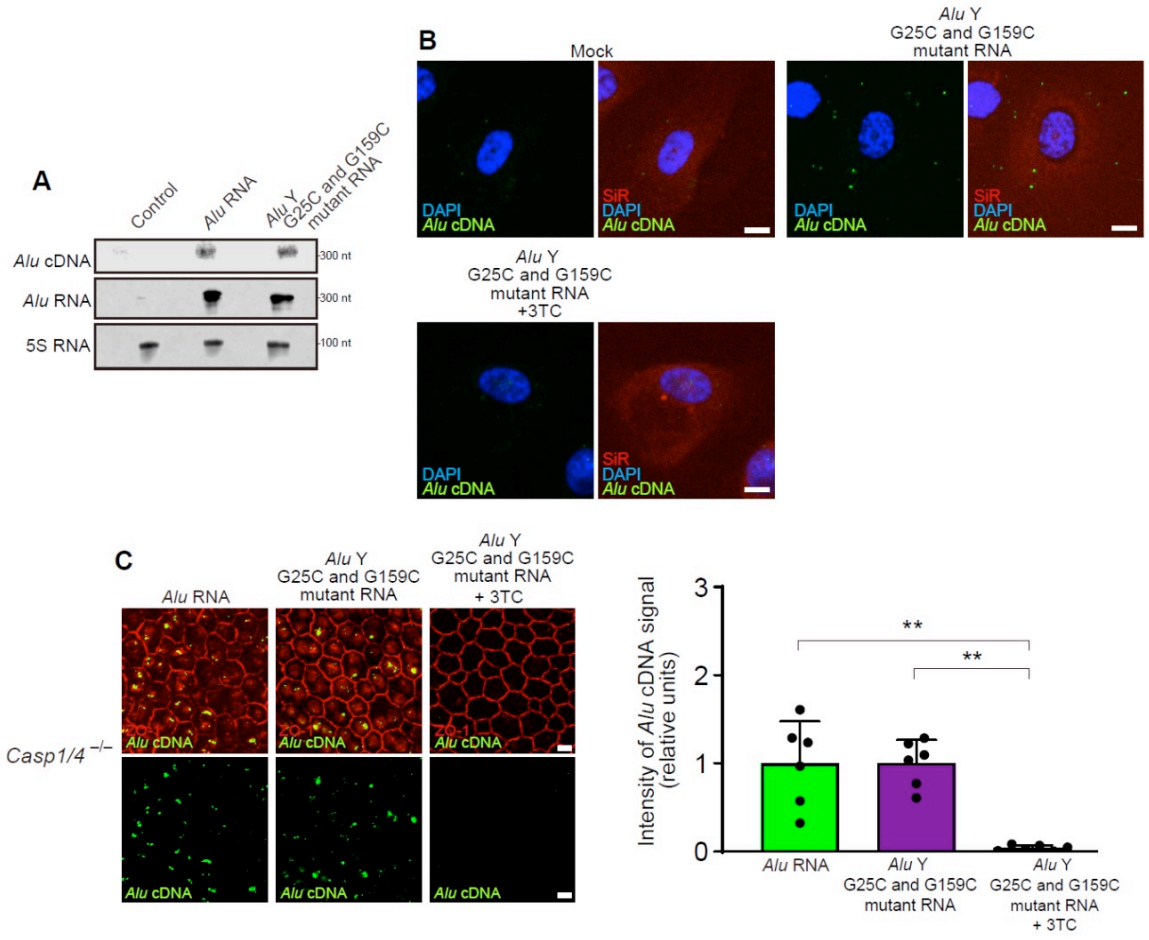


Fig. S14. Retrotransposition-deficient *Alu* G25C/G159C double mutant RNA induces RT-dependent cDNA synthesis. (A) Equator blotting of *Alu* cDNA in mouse embryonic carcinoma cells (F9) following transfection of in vitro transcribed *Alu* RNA or *Alu* G25C/G159C double mutant RNA. **(B)** In situ hybridization of *Alu* cDNA (green) in primary human RPE cells after treatment with in vitro transcribed *Alu* RNA or *Alu* G25C/G159C RNA following treatment with 3TC. DAPI (blue). Scale bars, 10 μm. (**C**) In situ hybridization to detect *Alu* cDNA (green) of RPE whole mounts from Casp1/4 dko mice after subretinal administration of in vitro transcribed *Alu* RNA or *Alu* G25C/159C double mutant RNA following treatment with 3TC. $n = 6$. Quantification of signal below. $^{**}P < 0.01$ by Mann-Whitney U test. Error bars show SEM.

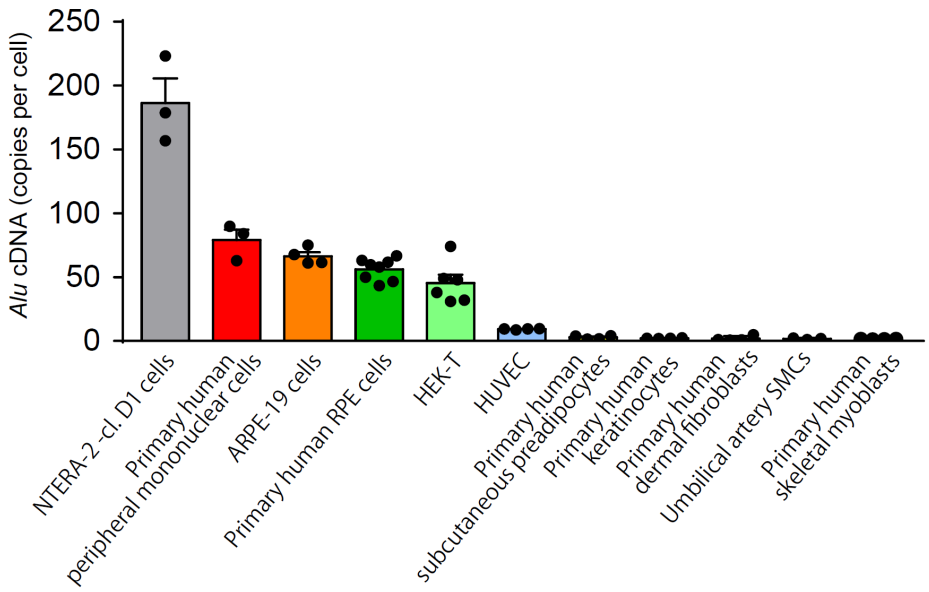
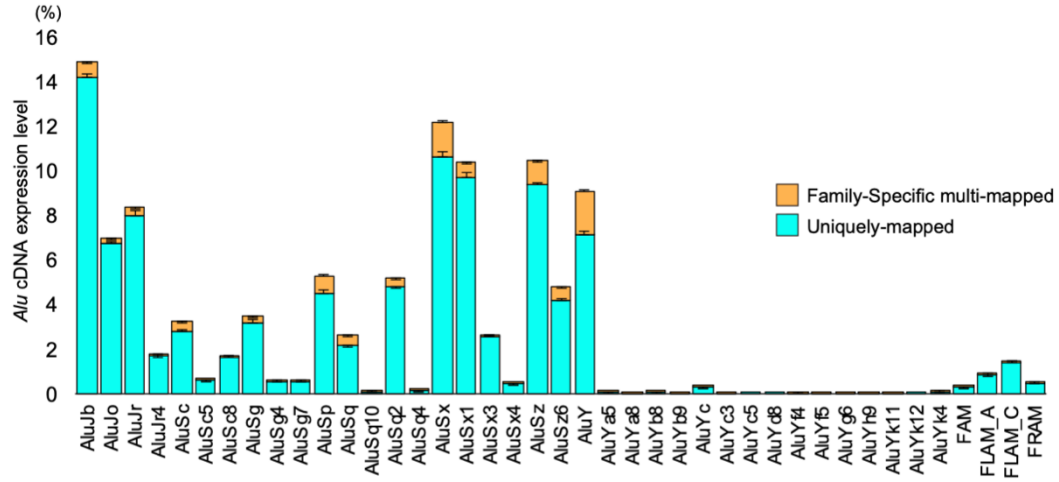


Fig. S15. Abundance of endogenous *Alu* cDNA in various human cell types. Copy number of *Alu* cDNA per cell, quantified in NTera2D cells, primary human peripheral blood mononuclear cells, ARPE-19 cells, primary human RPE cells, and human embryonic kidney-293-T cells (HEK-T), human umbilical vein endothelial cells (HUVEC), primary human subcutaneous pre-adipocytes, Primary human epidermal keratinocytes, primary human dermal fibroblasts, umbilical artery vascular smooth muscle cells (SMCs), and primary human skeletal myoblasts. The copy number of *Alu* cDNA was calculated using standard curves that were obtained using serial dilutions of the plasmids containing an *Alu*Y sequence. *Alu* cDNA copy number was normalized to cell number. n = 3–8. Error bars show SEM.

A



B

	***Alu* J & *Alu* S**	***Alu* Y**
Our study (*Alu* cDNA)	92%	8%
Zhang et al. 2019 (37) – Figure S3B and S6A (Pol III *Alu* RNA)	90%	10%
Conti et al. 2015 (38) – Table 2 (Pol III *Alu* RNA)	94%	6%
Oler et al. 2012 (39) – Table S3 (Pol III *Alu* RNA)	88%	12%

Fig. S16. *Alu* cDNA subfamily analysis from cytoplasm of primary human RPE cells. (A) Relative abundance of cDNAs derived from different *Alu* subfamilies in the cytoplasmic fraction of primary human RPE cells. The stacked bar plot shows the fractions of uniquely and multi-mapped (all alignments mapping within same subfamily) *Alu* read counts per subfamily. $n = 4$. Error bars show SEM. (**B**) Distribution of *Alu* cDNA and Pol III-derived *Alu* RNA by ancient (*Alu*J & *Alu*S) or young (*Alu*Y) subfamilies.

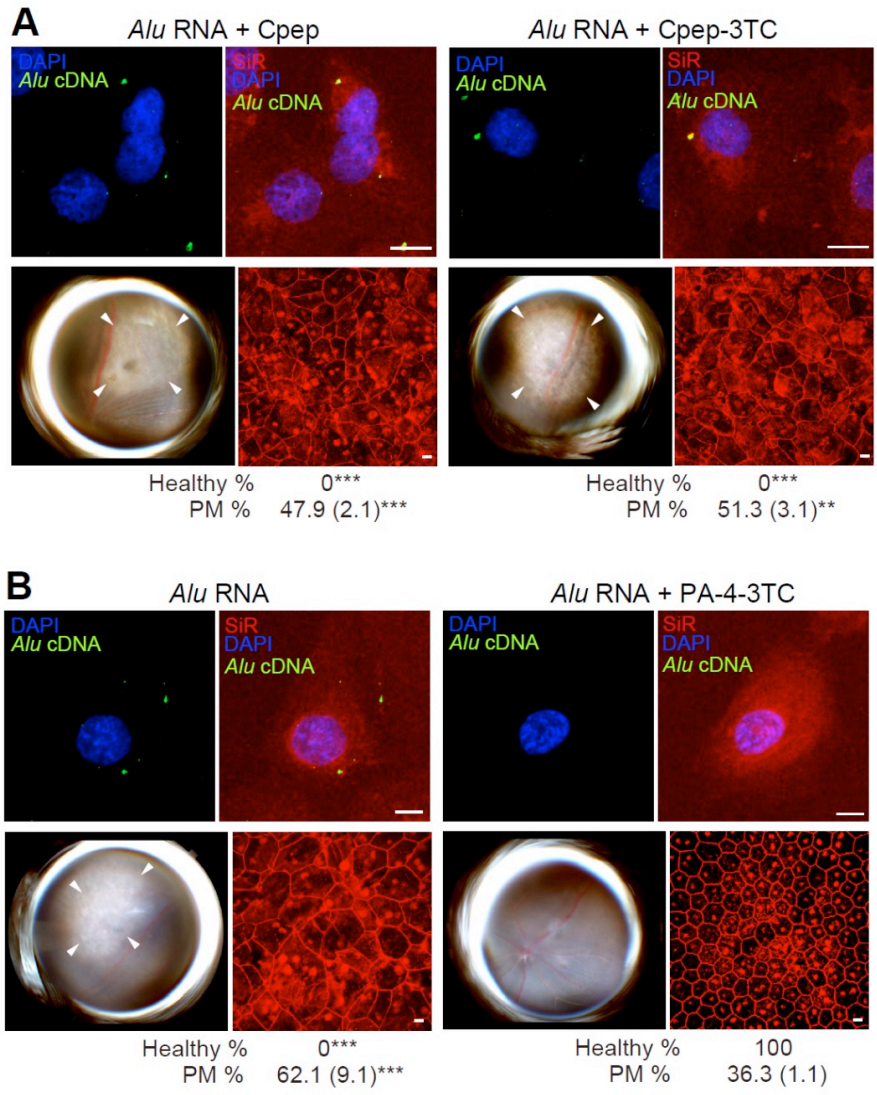


Fig. S17. *Alu* cDNA is synthesized in the cytoplasm. (A and **B)** Top row shows *Alu* cDNA formation monitored by in situ hybridization of primary human RPE cells transfected with *Alu* RNA. Green, *Alu* cDNA; Red, SiR-actin; Blue, DAPI. Cells were treated with Cpep-3TC (nuclear-targeting cyclic peptide-conjugated 3TC) or control peptide (Cpep) **(A)** or PA-4-3TC (cytoplasmic-targeting NRTI formulation) (**B**). Scale bar, 10 μm. Representative of $n = 6$. Bottom row shows pairs of fundus photographs (left) and corresponding representative RPE sheet micrographs (right) of WT mice following administration of *Alu* RNA and Cpep-3TC or control peptide (Cpep) **(A)** or PA-4-3TC (**B**). Scale bars, 10 μm. Arrowheads in fundus images denote the boundaries of RPE hypopigmentation. Binary and morphometric quantification of RPE degeneration are shown. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$, Fisher's exact test for binary; two-tailed t-test for morphometry. PM, polymegethism (mean (SEM)). $n = 6–12$.

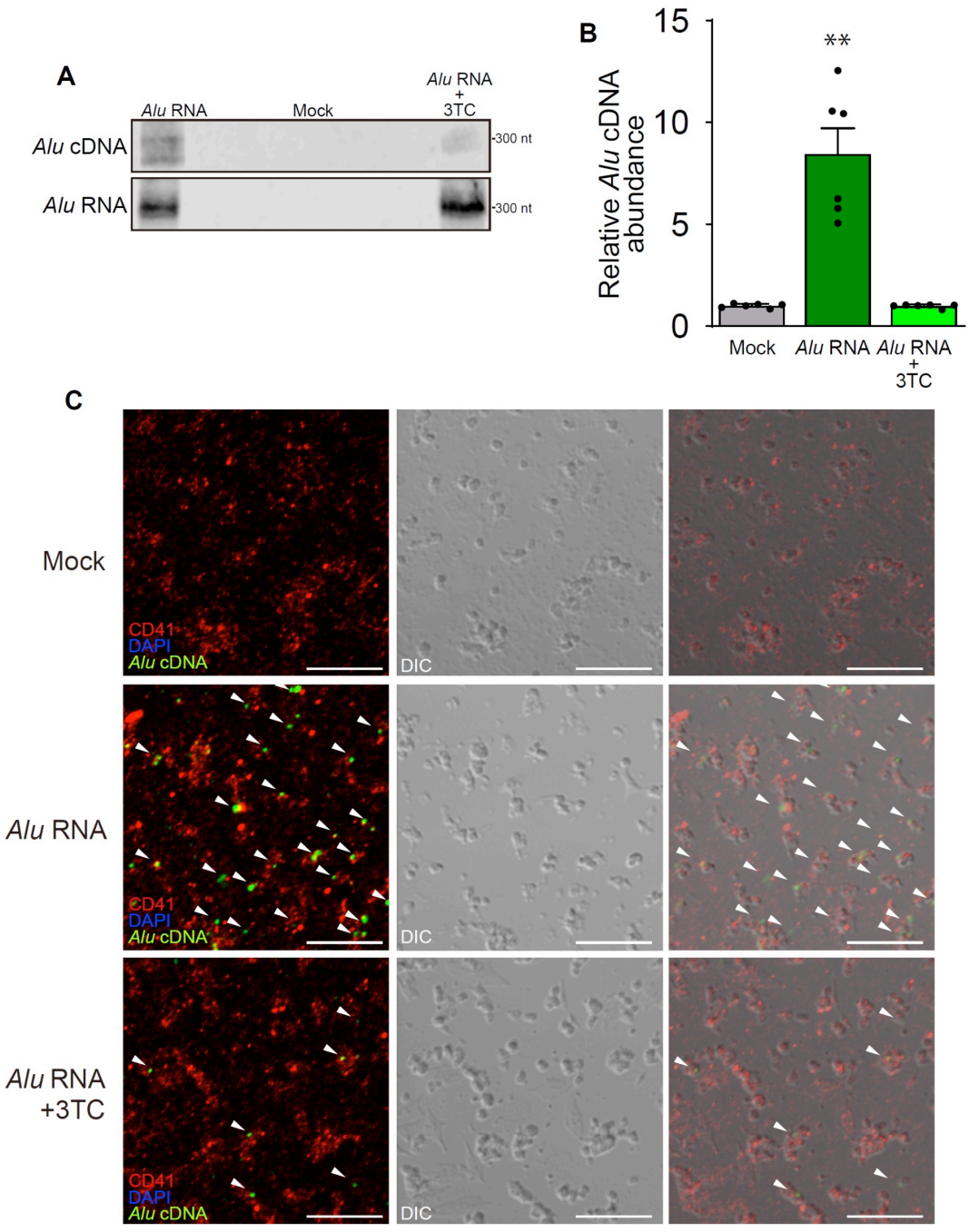


Fig. S18. *Alu* cDNA detection in mouse platelets transfected with *Alu* RNA. *Alu* cDNA quantification of wild-type (WT) mouse platelets after *Alu* RNA transfection in the presence or absence of 3TC treatment by **(A)** Equator blotting, **(B)** c-PCR, and **(C)** in situ hybridization. Arrowheads point to *Alu* cDNA signal (green). CD41, marker of mouse platelets (red), DAPI (blue). Scale bars, 10 μm. *$P < 0.05$; ** $P < 0.01$ by Mann-Whitney U test. DIC, differential interference contrast. Error bars show SEM.

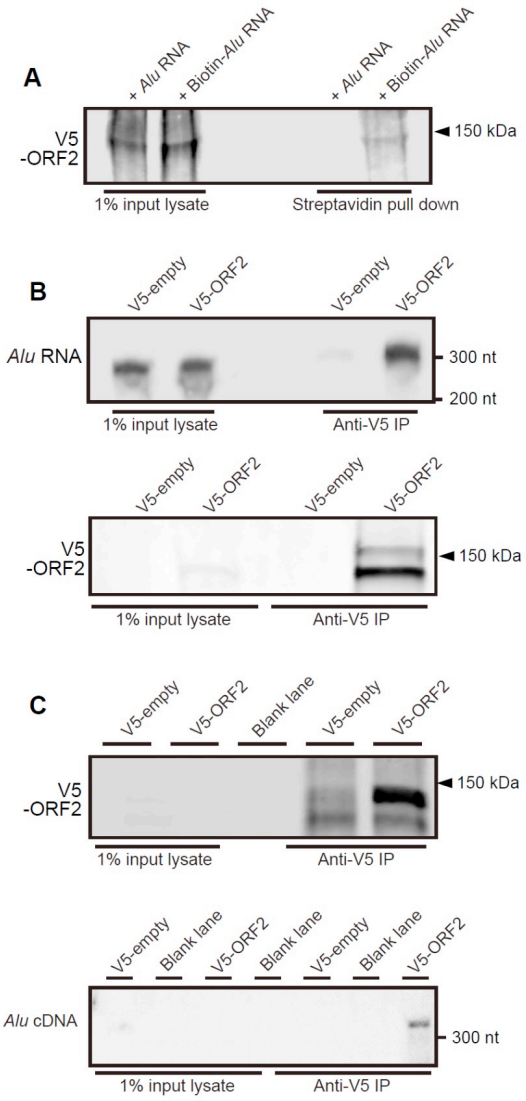


Fig. S19. Co-localization of L1 ORF2p, *Alu* RNA, and *Alu* cDNA. (A) Cytoplasmic fractions of RNaseH-deficient HeLa cells expressing V5-tagged L1 ORF2p (V5-ORF2) in the presence of transfected *Alu* RNA (either biotinylated or unlabeled) were subjected to streptavidin bead pull-down assay. Immunoblotting analysis using anti-V5 antibody of input and streptavidin-precipitated samples harboring biotin-labelled or unlabeled *Alu* RNA. **(B)** Cytoplasmic fractions of HeLa cells expressing V5-ORF2p or V5-empty plasmid in the presence of transfected biotinylated *Alu* RNA were subjected to anti-V5-immunoprecipitation. The precipitates were separated on polyacrylamide gels and analyzed for the presence of *Alu* RNA by Northern Blot analysis with an *Alu* RNA probe (upper panel) and immunoblot analysis using an anti-V5 antibody (lower panel). (**C**) Cytoplasmic fractions of RNaseH2 deficient MEF (*Rnaseh2b*$^{-/-}$) cells expressing V5-ORF2 or V5-empty in the presence of transfected *Alu* RNA were subjected to anti-V5-immunoprecipitation. The precipitates were subjected to immunoblot analysis using anti-V5 antibody (upper panel) and equator blotting detection of the *Alu* cDNA (lower panel).

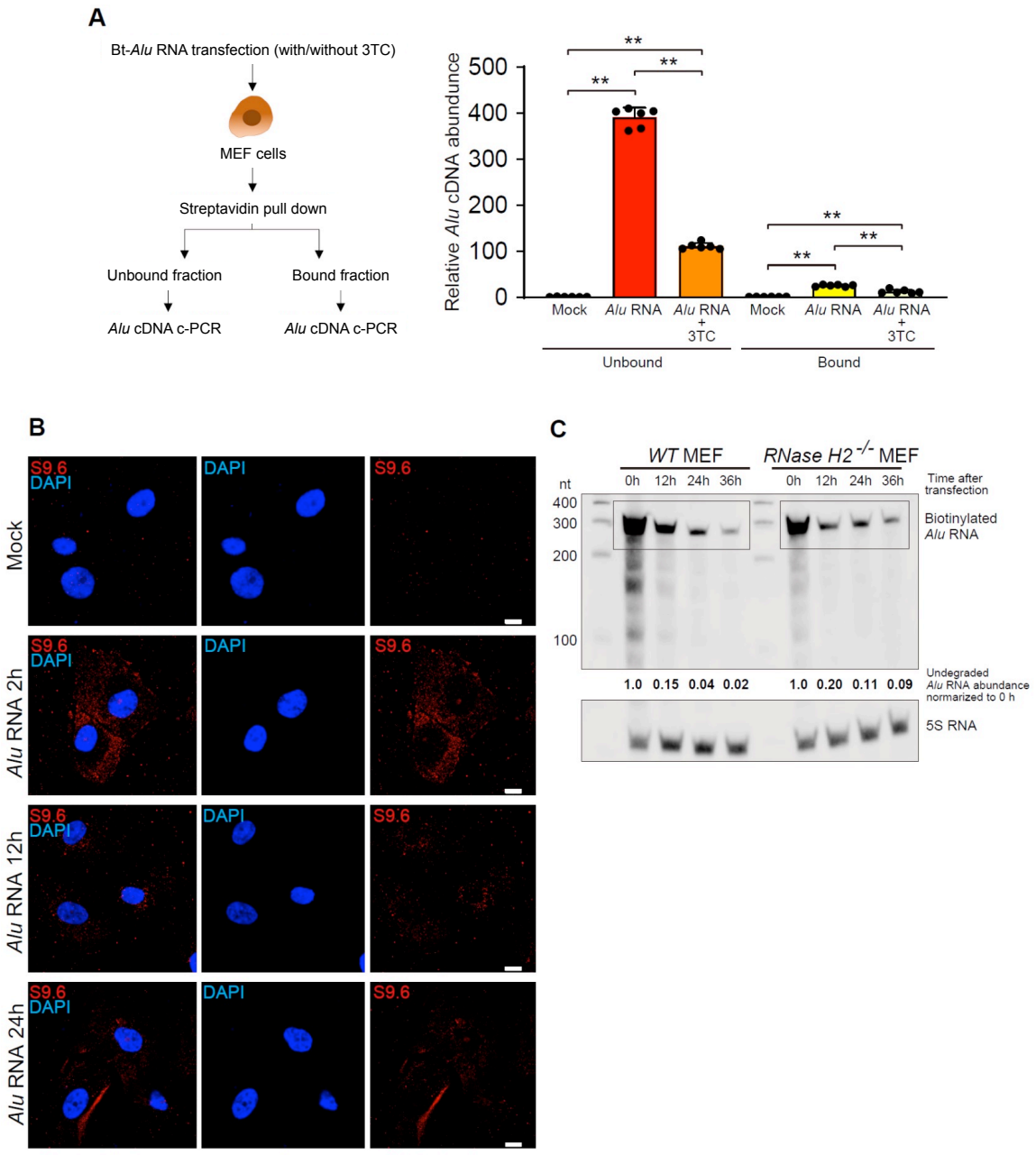


Fig. S20. RNA-DNA hybrids following *Alu* RNA transfection. (**A**) *Alu* RNA/cDNA hybrid assay to monitor association of *Alu* RNA and cDNA**.** Cytoplasmic fractions of MEF cells transfected with biotinylated *Alu* RNA (with or without 3TC) were subjected to streptavidin pull-down. Bound and unbound fractions from streptavidin pulldown were analyzed by *Alu* real-time PCR. ** $P < 0.01$ by Mann-Whitney U test. Error bars show SEM. **(B)** Visualization of cytosolic RNA-DNA hybrids in primary human RPE cells after *Alu* RNA transfection for the indicated times by S9.6 antibody staining (red). DAPI (blue). Scale bar, 10 μm. (**C**) Polyacrylamide gel-separated biotinylated RNA from control (*Rnaseh2b*$^{+/+}$) MEF and RNaseH2-deficient (*Rnaseh2b*$^{-/-}$) MEF cells at the indicated time points after biotinylated *Alu* RNA transfection. *Rnaseh2b*$^{-/-}$ and *Rnaseh2b*$^{+/+}$ control MEFs on a C57BL/6 p53$^{-/-}$ background.

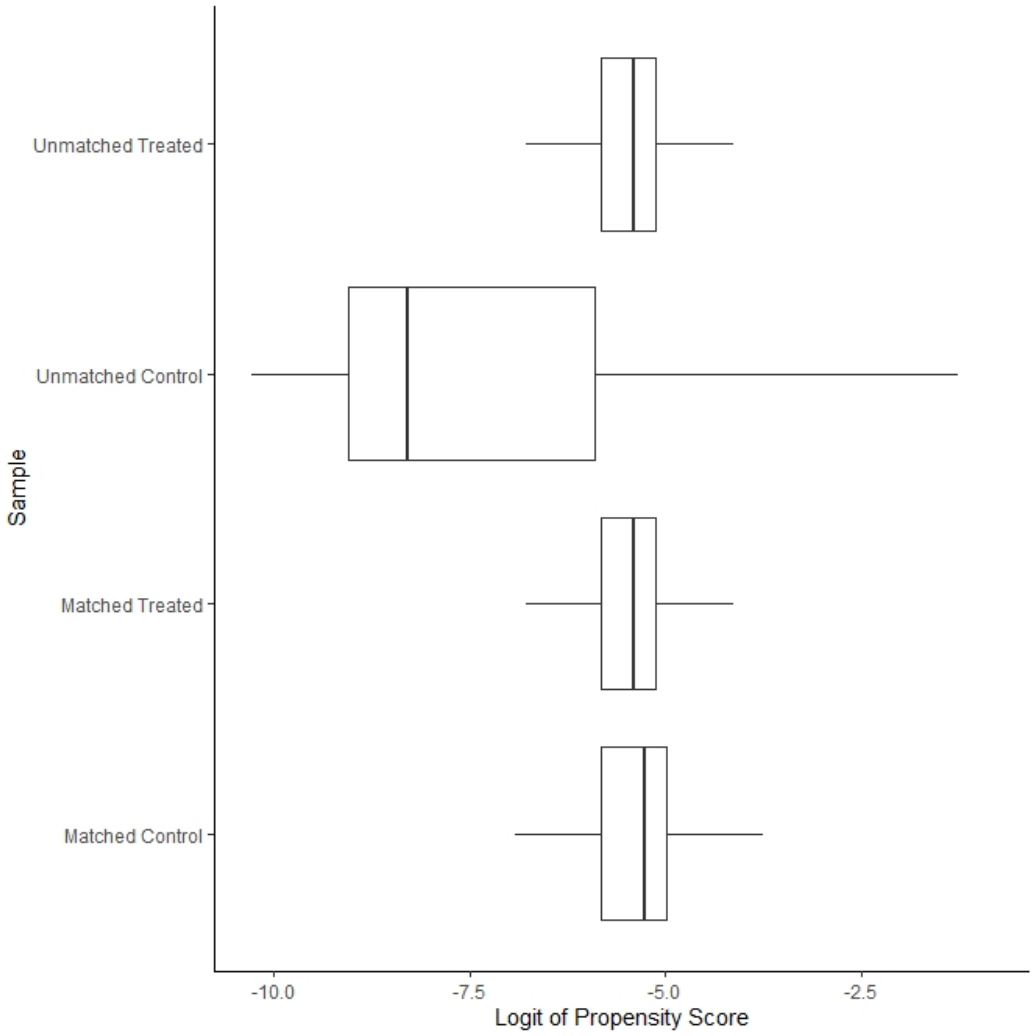


Fig. S21. Distribution of the Logit of the Propensity Score – Truven Marketscan Database. Box plots for the logit of the propensity score show good balance for the matched observations. The box represents the 1st quartile (left side), median (vertical line inside the box), and the 3rd quartile (right side). Lines extending on either side of the box are 1.5-times the interquartile range.

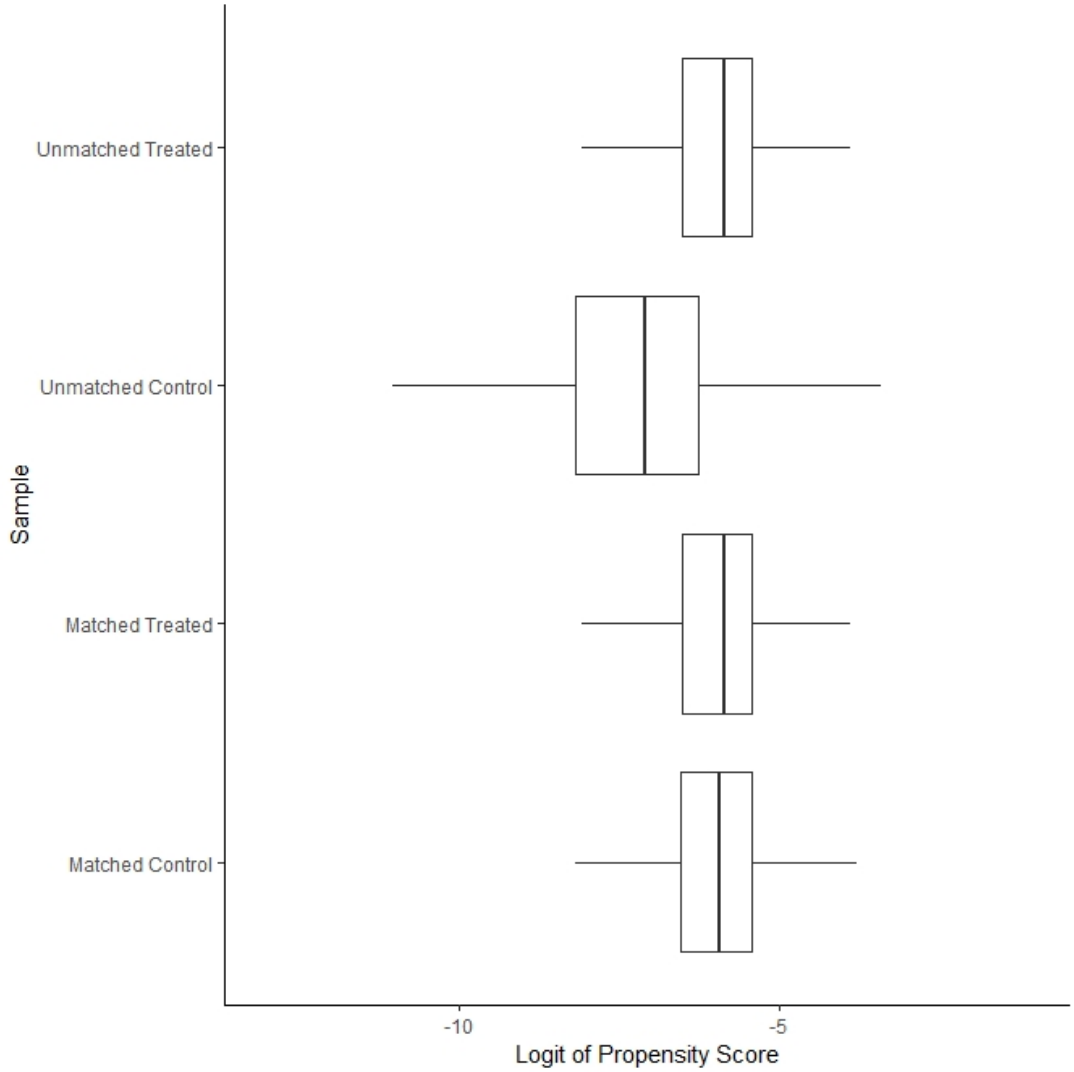


Fig. S22. Distribution of the Logit of the Propensity Score – Veterans Health Administration Database. Box plots for the logit of the propensity score show good balance for the matched observations. The box represents the 1st quartile (left side), median (vertical line inside the box), and the 3rd quartile (right side). Lines extending on either side of the box are 1.5-times the interquartile range.

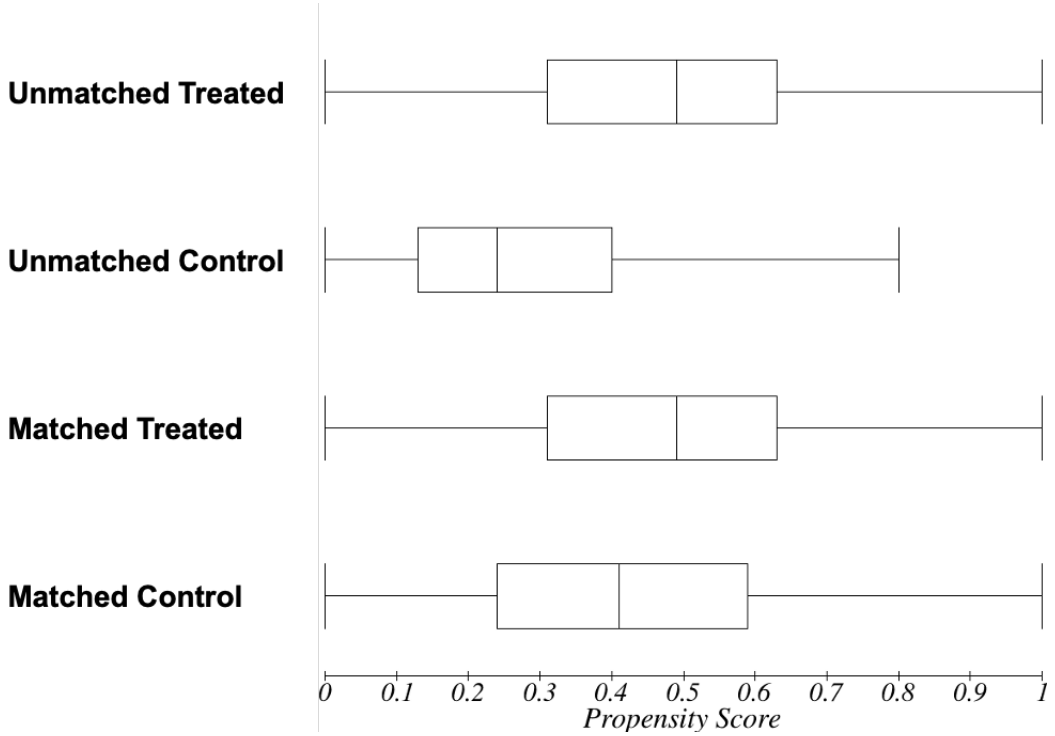


Fig. S23. Distribution of the Propensity Score – PearlDiver Humana Database. Box plots for the propensity score show good balance for the matched observations. The box represents the 1st quartile (left side), median (vertical line inside the box), and the 3rd quartile (right side). Lines extending on either side of the box are 1.5-times the interquartile range.

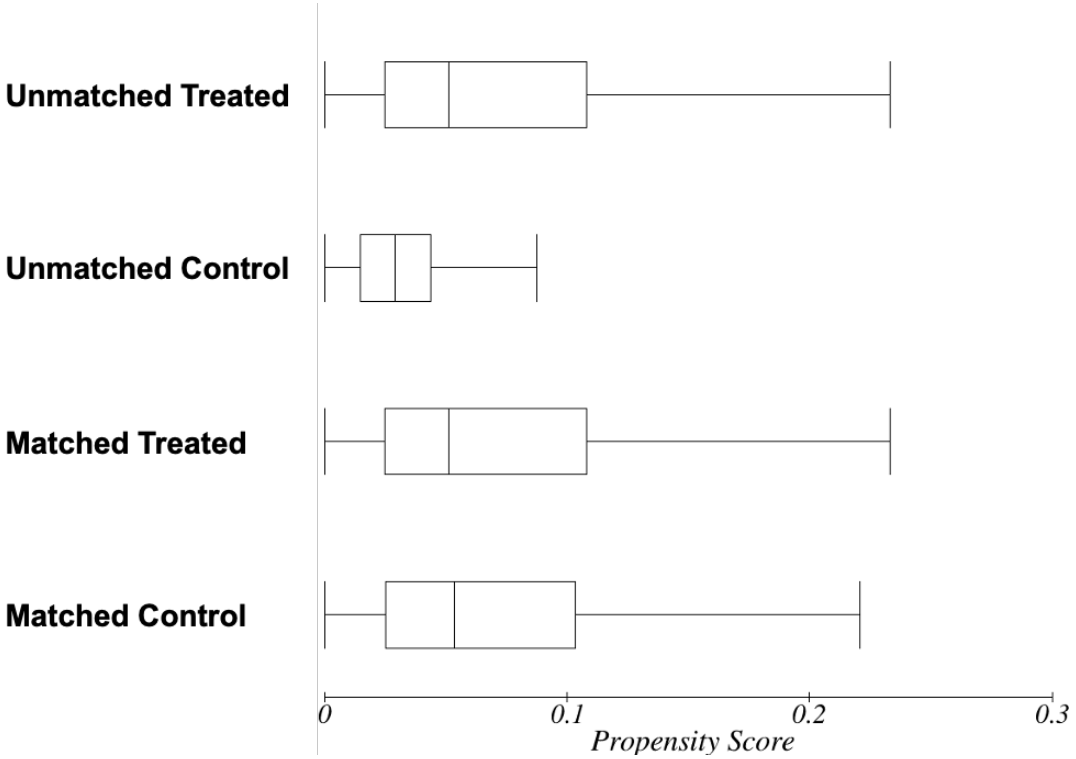


Fig. S24. Distribution of the Propensity Score – PearlDiver Mariner Database. Box plots for the propensity score show good balance for the matched observations. The box represents the 1st quartile (left side), median (vertical line inside the box), and the 3rd quartile (right side). Lines extending on either side of the box are 1.5-times the interquartile range.

Table S1. Incident atrophic AMD among HIV-negative persons.

Database	Users	Cases	Atrophic AMD rate
Truven			0.95%
NRTIs			
Never user	21,328,589	202,745	0.95%
Ever user	902	2	0.22%
Veterans			6.78%
NRTIs			
Never user	1,628,777	110,463	6.78%
Ever user	814	7	0.86%
Humana			6.68%
NRTIs			
Never user	5,620,092	375,649	6.68%
Ever user	5,761	85	1.48%
Mariner			1.93%
NRTIs			
Never user	4,927,160	95,027	1.93%
Ever user	11,499	53	0.46%
Pooled			2.34%
NRTIs			
Never user	33,504,618	783,884	2.34%
Ever user	18,976	147	0.77%

Incident atrophic AMD cases were less frequent among HIV-negative persons aged 50 years or older and exposed to nucleoside reverse transcriptase inhibitors (NRTIs) for pre-exposure prophylaxis compared with HIV-negative persons aged 50 years or older and never exposed to NRTIs. Truven Health MarketScan Commercial Claims and Encounters & Medicare Supplemental (100% sample), U.S. Veterans Health Administration (30% random sample for NRTI never user and 100% sample for NRTI user), PearlDiver Mariner (100% sample), and PearlDiver Humana (100% sample) databases.

Table S2. Baseline characteristics of Truven Marketscan Database.

Variable		No record of NRTI N=514,364	NRTI Exposure N=902	P value*
Index Age – Mean (SD)		56.1 (3.93)	54.2 (3.58)	<0.001
Sex	**Female**	55.7%	4.8%	<0.001
	Male	44.3%	95.2%	
Charlson comorbidity index Mean (SD)		0.16 (0.63)	0.39 (1.02)	<0.001
Smoking		0.7%	3.7%	<0.001
Obesity		1.0%	6.6%	<0.001

Baseline characteristics of all NRTI users and a random sample of NRTI non-users in the Truven Database comprising the study population analyzed using a Bayesian proportional hazards Cox regression model. *P-values for continuous variables are from Student t-tests and categorical from chi-square (χ^2) tests. All statistical tests are two-sided.

Table S3. Baseline characteristics of Veterans Database.

Variable		No record of NRTI N=578,767	NRTI Exposure N=814	P value*
Index age – Mean (SD)		66.5 (9.99)	58.3 (7.1)	<0.001
Race	**Black**	9.3%	17.1%	<0.001
	Other	20.3%	7.1%	
	White	70.4%	75.8%	
Sex	**Female**	2.6%	3.2%	0.355
	Male	97.4%	96.8%	
Charlson comorbidity index Mean (SD)		0.28 (0.77)	0.69 (1.2)	<0.001
Smoking		28.2%	34.2%	<0.001
BMI	**<18.5**	1.2%	0.4%	<0.001
	18.5-24.9	20.9%	19.0%	
	25-29.9	38.8%	37.7%	
	30+	35.8%	42.0%	
	Missing	3.2%	0.9%	

Baseline characteristics of all NRTI users and a random sample of NRTI non-users in the Veterans Database comprising the study population analyzed using a Bayesian proportional hazards Cox regression model. *P-values for continuous variables are from Student t-tests and categorical from chi-square (χ^2) tests. All statistical tests are two-sided.

Table S4. Baseline characteristics of PearlDiver Humana Database.

Variable		No record of NRTI N=5,620,092	NRTI Exposure N=5,761	P value
Index Age		68.2 (13.8)	60.2 (8.09)	<0.001
Race	Black	9.7%	25.2%	<0.001
	Other	29.2%	9.0%	
	White	61.1%	65.8%	
Sex	Female	57.0%	33.3%	<0.001
	Male	43.0%	66.7%	
Charlson comorbidity index Mean (SD)		0.22 (0.66)	0.78 (1.96)	<0.001
Smoking		13.1%	17.0%	<0.001
BMI 30–40		15.0%	10.6%	<0.001

Baseline characteristics of all NRTI users and all NRTI non-users in the PearlDiver Humana Database comprising the study population analyzed using a Cox proportional hazards regression model. *P-values for continuous variables are from Student t-tests and categorical from chi-square (χ^2) tests. All statistical tests are two-sided.

Table S5. Baseline characteristics of PearlDiver Mariner Database.

Variable		No record of NRTI N=479,064	NRTI Exposure N=11,499	P value*
Index Age		64.3 (8.28)	56.7 (6.39)	<0.001
Sex	**Female**	58.6%	32.5%	
	Male	41.4%	67.5%	<0.001
Charlson comorbidity index Mean (SD)		0.87 (1.67)	1.97 (3.03)	<0.001
Smoking		11.2%	23.2%	<0.001
BMI 30–40		10.8%	9.1%	<0.001

Baseline characteristics of all NRTI users and a random sample of NRTI non-users in the PearlDiver Mariner Database comprising the study population analyzed using a Cox proportional hazards regression model. *P-values for continuous variables are from Student t-tests and categorical from chi-square (χ^2) tests. All statistical tests are two-sided.

Table S6. Sensitivity analysis of Bayesian priors for VA & Truven – Propensity Unmatched Analysis.

	Priors		
	M = 20, r_0 = 0.1	**M = 20, r_0 = 1**	**M = 20, r_0 = 100**
Database	**Adjusted HR (95% CI)**	**Adjusted HR (95% CI)**	**Adjusted HR (95% CI)**
Truven	0.477 (0.461-0.495)	0.421 (0.376-0.501)	0.453 (0.420-0.482)
Veterans	0.760 (0.651-0.870)	0.838 (0.709-0.965)	0.784 (0.708-0.834)

Sensitivity analyses for the prior specification changing the hyperparameter to M = 20 and using r_0 values of 0.1, 1 and 100. M is the number of pieces of the hazard function. r_0 is the weight given to the exponential distribution underlying the baseline hazard function. When r_0 = 0.1, very little weight is given to the centering exponential distribution, whereas r_0 = 1000 essentially fits an exponential regression model. We present estimated mean, and 95% lower and upper values of the adjusted hazard ratio (HR).

Table S7. Hazards of Incident Appendicitis and Hernia for NRTI use.

	Appendicitis	Hernia
Database	**Adjusted HR (95% CI)**	**Adjusted HR (95% CI)**
Truven	1.771 (0.981-3.198)	0.903 (0.743-1.096)
Veterans	3.065 (0.766-12.27)	0.811 (0.528-1.243)
Mariner	0.986 (0.748-1.299)	0.974 (0.920-1.031)

Adjusted hazard ratios (HR) and 95% confidence intervals (CI) for incident appendicitis or incident hernia are presented for database.

Table S8. Baseline characteristics of propensity score matched population – Truven Database

Variable		No record of NRTI N=902	NRTI Exposure N=902	P value*
Index age – Mean (SD)		53.9 (3.62)	54.2 (3.58)	0.049
Sex	**Female**	4.0%	4.8%	0.42
	Male	96.0%	95.2%	
Charlson comorbidity		0.43 (1.32)	0.39 (1.02)	0.20
Smoking		3.4%	3.7%	0.80
Obesity		7.7%	6.6%	0.41
NRTI treatment duration Days – Mean (SD)		0 (0)	413.5 (354)	<0.001
Follow-up Days – Mean (SD)		724.6 (478.1)	578.6 (452.2)	<0.001

*P-values for continuous variables are from t-tests and categorical from chi-square (χ^2) tests. All statistical tests are two-sided.

Table S9. Baseline characteristics of propensity score matched population – Veterans Database

Variable		No record of NRTI N=814	NRTI Exposure N=814	P value*
Index age		58.2 (7.31)	58.3 (7.1)	0.75
Race	**Black**	18.9%	17.1%	0.61
	Other	7.3%	7.1%	
	White	73.8%	75.8%	
Sex	**Female**	3.4%	3.2%	0.89
	Male	96.6%	96.8%	
Charlson comorbidity		0.62 (1.48)	0.7 (1.22)	0.28
Smoking		30.7%	34.2%	0.15
BMI	**<18.5**	0.2%	0.4%	
	18.5-24.9	17.7%	19.0%	0.68
	25-29.9	39.4%	37.7%	
	30+	41.2%	42.0%	
	Missing	1.5%	0.9%	
NRTI treatment duration, Days – Mean (SD)		0 (0)	502.5 (417.1)	<0.001
Follow-up Days – Mean (SD)		667.8 (590.6)	707.23 (487.6)	0.14

*P-values for continuous variables are from t-tests and categorical from chi-square (χ^2) tests. All statistical tests are two-sided.

Table S10. Baseline characteristics of propensity score matched population – PearlDiver Humana Database.

Variable		No record of NRTI N=5,761	NRTI Exposure N=5,761	P value*
Index age – Mean (SD)		62.8 (8.30)	60.2 (8.09)	<0.001
Sex	**Female**	42.1%	33.3%	<0.001
	Male	57.9%	66.7%	
Charlson comorbidity index, Mean (SD)		0.76 (0.27)	0.78 (1.96)	0.44
Smoking		13.2%	17.0%	<0.001
BMI 30–40		11.4%	10.6%	0.17
NRTI treatment duration Days – Mean (SD)		0 (0)	722.2 (821.5)	<0.001
Follow-up Days – Mean (SD)		1,586.2 (1,110.1)	1,270.8 (1,100.7)	<0.001

*P-values for continuous variables are from t-tests and categorical from chi-square (χ^2) tests. All statistical tests are two-sided.

Table S11. Baseline characteristics of propensity score matched population – PearlDiver Mariner Database.

Variable		No record of NRTI N=11,499	NRTI Exposure N=11,499	P value*
Index Age – Mean (SD)		57.1 (6.42)	56.7 (6.39)	<0.001
				0.80
Sex	**Female**	32.4%	32.5%	
	Male	67.6%	67.5%	
Charlson comorbidity index Mean (SD)		2.17 (3.50)	1.97 (3.03)	<0.001
Smoking		22.7%	23.2%	0.36
BMI 30–40		9.3%	9.1%	0.62
NRTI treatment duration, Days – Mean (SD)		0 (0)	1,164.9 (950.4)	<0.001
Follow-up Days – Mean (SD)		2,330.0 (803.7)	2,631.3 (613.2)	<0.001

*P-values for continuous variables are from t-tests and categorical from chi-square (χ^2) tests. All statistical tests are two-sided.

Table S12. Sensitivity analysis of Bayesian priors for VA & Truven – Propensity Matched Analysis.

	Priors		
	M = 20, r_0 = 0.1	**M = 20, r_0 = 1**	**M = 20, r_0 = 100**
Database	**Adjusted HR (95% CI)**	**Adjusted HR (95% CI)**	**Adjusted HR (95% CI)**
Truven	0.762 (0.754-0.772)	0.786 (0.743-1.096)	0.772 (0.738-0.797)
Veterans	0.947 (0.941-0.953)	0.956 (0.949-0.962)	0.946 (0.939-0.950)

Sensitivity analyses for the prior specification changing the hyperparameter to M = 20 and using r_0 values of 0.1, 1 and 100. M is the number of pieces of the hazard function. r_0 is the weight given to the exponential distribution underlying the baseline hazard function. When r_0 = 0.1, very little weight is given to the centering exponential distribution, whereas r_0 = 1000 essentially fits an exponential regression model. We present estimated mean, and 95% lower and upper values of the adjusted hazard ratio (HR).

SI References

1. J. Ambati, B. Fowler, Compositions and Methods for Treating Retinal Degradation. *Ophthalmol. Vis. Sci. Fac. Patents* (2016) (October 28, 2020).
2. D. Mandal, A. Nasrolahi Shirazi, K. Parang, Cell-penetrating homochiral cyclic peptides as nuclear-targeting molecular transporters. *Angew. Chemie - Int. Ed.* **50**, 9633–9637 (2011).
3. A. Nasrolahi Shirazi, *et al.*, Peptide amphiphile containing arginine and fatty acyl chains as molecular transporters. *Mol. Pharm.* **10**, 4717–4727 (2013).
4. N. Kerur, *et al.*, TLR-independent and P2X7-dependent signaling mediate Alu RNA-induced NLRP3 inflammasome activation in geographic atrophy. *Investig. Ophthalmol. Vis. Sci.* **54**, 7395–7401 (2013).
5. H. Kaneko, *et al.*, DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. *Nature* **471**, 325–332 (2011).
6. V. Tarallo, *et al.*, DICER1 loss and Alu RNA induce age-related macular degeneration via the NLRP3 inflammasome and MyD88. *Cell* **149**, 847–859 (2012).
7. B. J. Fowler, *et al.*, Nucleoside reverse transcriptase inhibitors possess intrinsic anti-inflammatory activity. *Science.* **346**, 1000–1003 (2014).
8. N. Kerur, *et al.*, CGAS drives noncanonical-inflammasome activation in age-related macular degeneration. *Nat. Med.* **24**, 50–61 (2018).
9. Y. Kim, *et al.*, DICER1/Alu RNA dysmetabolism induces caspase-8-mediated cell death in age-related macular degeneration. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1682–1687 (2014).
10. J. Huang, *et al.*, Comparison of noncontact specular and confocal microscopy for evaluation of corneal endothelium. *Eye Contact Lens* **44**, S144–S150 (2018).
11. S. Dridi, *et al.*, ERK1/2 activation is a therapeutic target in age-related macular degeneration. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 13781–13786 (2012).
12. T. Ach, *et al.*, Lipofuscin redistribution and loss accompanied by cytoskeletal stress in retinal pigment epithelium of eyes with age-related macular degeneration. *Investig. Ophthalmol. Vis. Sci.* **56**, 3242–3252 (2015).
13. H. E. Grossniklaus, J. M. Nickerson, H. F. Edelhauser, L. A. M. K. Bergman, L. Berglin, Anatomic alterations in aging and age-related diseases of the eye. *Investig. Ophthalmol. Vis. Sci.* **54**, ORSF23-7 (2012).
14. I. S. Tarau, A. Berlin, C. A. Curcio, T. Ach, The Cytoskeleton of the Retinal Pigment Epithelium: from Normal Aging to Age-Related Macular Degeneration. *Int. J. Mol. Sci.* **20** (2019).
15. M. S. Taylor, *et al.*, Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. *Cell* **155**, 1034–1048 (2013).
16. A. Kirilyuk, *et al.*, Functional endogenous LINE-1 retrotransposons are expressed and mobilized in rat chloroleukemia cells. *Nucleic Acids Res.* **36**, 648–665 (2008).
17. A. Troegeler, *et al.*, An efficient siRNA-mediated gene silencing in primary human monocytes, dendritic cells and macrophages. *Immunol. Cell Biol.* **92**, 699–708 (2014).
18. W. man Liu, W. ming Chu, P. V. Choudary, C. W. Schmid, Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res.* **23**, 1758–1765 (1995).
19. A. J. Doucet, *et al.*, Characterization of LINE-1 Ribonucleoprotein Particles. *PLoS Genet.* **6**, e1001150 (2010).
20. J. Hoebeeck, *et al.*, Rapid detection of VHL exon deletions using real-time quantitative PCR. *Lab. Investig.* **85**, 24–33 (2005).
21. B. D'haene, J. Vandesompele, J. Hellemans, Accurate and objective copy number profiling using real-time quantitative PCR. *Methods* **50**, 262–270 (2010).
22. W. Wei, *et al.*, Human L1 Retrotransposition: cisPreference versus trans Complementation. *Mol. Cell. Biol.* **21**, 1429–1439 (2001).
23. K. J. Mackenzie, *et al.*, Ribonuclease H2 mutations induce a cGAS / STING -dependent innate immune response . *EMBO J.* **35**, 831–844 (2016).
24. D. Bhavanasi, S. Kim, L. E. Goldfinger, S. P. Kunapuli, Protein kinase Cδ mediates the activation of protein kinase D2 in platelets. *Biochem. Pharmacol.* **82**, 720–727 (2011).
25. K. Wang, *et al.*, MapSplice: Accurate mapping of RNA-seq reads for splice junction

discovery. *Nucleic Acids Res.* **38** (2010).
26. A. Dobin, *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
27. Y. Liao, G. K. Smyth, W. Shi, FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
28. E. Vittinghoff, C. E. McCulloch, Relaxing the rule of ten events per variable in logistic and cox regression. *Am. J. Epidemiol.* **165**, 710–718 (2007).
29. H. Zhou, T. Hanson, J. Zhang, Spbayessurv: Fitting bayesian spatial survival models using R. *J. Stat. Softw.* **92**, 1–33 (2020).
30. J. P. T. Higgins, S. G. Thompson, Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
31. C. Anello, J. L. Fleiss, Exploratory or analytic meta-analysis: Should we distinguish between them? *J. Clin. Epidemiol.* **48**, 109–116 (1995).
32. R. DerSimonian, N. Laird, Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177–188 (1986).
33. J. Lau, J. P. A. Ioannidis, C. H. Schmid, Summing up evidence: One answer is not always enough. *Lancet* **351**, 123–127 (1998).
34. D. J. Spiegelhalter, K. R. Abrams, J. P. Myles, *Bayesian approaches to clinical trials and healthcare evaluation*. Statistics in practice (John Wiley & Sons, Chichester; Hoboken, NJ, 2004), pp.xiv, 391 p.
35. N. G. Polson, J. G. Scott, On the half-Cauchy prior for a global scale parameter. *Bayesian Anal.* **7**, 887-902 (2012).
36. A. Gelman, Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.* **1**, 515-534 (2006).
37. X. O. Zhang, T. R. Gingeras, Z. Weng, Genome-wide analysis of polymerase III-transcribed Alu elements suggests cell-type-specific enhancer function. *Genome Res* **29**, 1402-1414 (2019).
38. A. Conti *et al.*, Identification of RNA polymerase III-transcribed Alu loci by computational screening of RNA-Seq data. *Nucleic Acids Res* **43**, 817-835 (2015).
39. A. J. Oler *et al.*, Alu expression in human cell lines and their retrotranspositional potential. *Mob DNA* **3**, 11 (2012).