PNAS RESEARCH ARTICLE | GENETICS OPEN ACCESS

CASP8 intronic expansion identified by poly-glycine-arginine pathology increases Alzheimer's disease risk

Lien Nguyen[a,b,c,1], Ramadan Ajredini[a,b], Shu Guo[a,b], Lisa E. L. Romano[a,b], Rodrigo F. Tomas[a,b], Logan R. Bell[a,b], Paul T. Ranum[d,e], Tao Zu[a,b], Monica Bañez Coronel[a,b], Chase P. Kelley[a,b], Javier Redding-Ochoa[f], Evangelos Nizamis[g], Alexandra Melloni[h], Theresa R. Connors[h], Angelica Gaona[h], Kiruphagaran Thangaraju[a,b], Olga Pletnikova[f], H. Brent Clark[i], Beverly L. Davidson[d,e], Anthony T. Yachnis[a,j], Todd E. Golde[k,l,m], XiangYang Lou[n], Eric T. Wang[a,b,c], Alan E. Renton[o,p], Alison Goate[o,p], Paul N. Valdmanis[g], Stefan Prokop[a,j,k,m,q], Juan C. Troncoso[f], Bradley T. Hyman[h], and Laura P. W. Ranum[a,b,c,m,q,r,1]

Affiliations are included on p. 11.

Edited by Jack Griffith, The University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC; received August 23, 2024; accepted December 6, 2024

Alzheimer's disease (AD) affects more than 10% of the population ≥65 y of age, but the underlying biological risks of most AD cases are unclear. We show anti-poly-glycine-arginine (a-polyGR) positive aggregates frequently accumulate in sporadic AD autopsy brains (45/80 cases). We hypothesize that these aggregates are caused by one or more polyGR-encoding repeat expansion mutations. We developed a CRISPR/deactivated-Cas9 enrichment strategy to identify candidate GR-encoding repeat expansion mutations directly from genomic DNA isolated from a-polyGR(+) AD cases. Using this approach, we isolated an interrupted $(GGGAGA)_n$ intronic expansion within a SINE-VNTR-Alu element in *CASP8* (*CASP8*-GGGAGA^{EXP}). Immunostaining using a-polyGR and locus-specific C-terminal antibodies demonstrate that the *CASP8*-GGGAGA^{EXP} expresses hybrid poly(GR)n(GE)n(RE)n proteins that accumulate in *CASP8*-GGGAGA^{EXP}(+) AD brains. In cells, expression of *CASP8*-GGGAGA^{EXP} minigenes leads to increased p-Tau (Ser202/Thr205) levels. Consistent with other types of repeat-associated non-AUG (RAN) proteins, poly(GR)n(GE)n(RE)n protein levels are increased by stress. Additionally, levels of these stress-induced proteins are reduced by metformin. Association studies show specific aggregate promoting interrupted *CASP8*-GGGAGA^{EXP} sequence variants found in ~3.6% of controls and 7.5% AD cases increase AD risk [*CASP8*-GGGAGA-AD-R1; OR 2.2, 95% CI (1.5185 to 3.1896), $P = 3.1 \times 10^{-5}$]. Cells transfected with a high-risk *CASP8*-GGGAGA-AD-R1 variant show increased toxicity and increased levels of poly(GR)n(GE)n(RE)n aggregates. Taken together, these data identify polyGR(+) aggregates as a frequent and unexpected type of brain pathology in AD and *CASP8*-GGGAGA-AD-R1 alleles as a relatively common AD risk factor. Taken together, these data support a model in which *CASP8*-GGGAGA^{EXP} alleles combined with stress increase AD risk.

microsatellite expansion mutations | Alzheimer's disease | repeat associated non-AUG (RAN) translation | protein aggregates | RAN proteins

Significance

Alzheimer's disease (AD) affects >10% of older adults but the causes are not well understood. We found proteins containing repeated glycine and arginine (polyGR) form aggregates that frequently accumulate in AD brains. We used polyGR-encoding guide RNAs and deactivated-Cas9 to isolate an intronic interrupted (GGGAGA)n expansion in *CASP8,* which produces hybrid poly(GR)n(GE)n(RE)n proteins that accumulate in *CASP8*-GGGAGA^{EXP}(+) AD brains. In cells, *CASP8*-GGGAGA^{EXP} increase p-Tau and poly(GR)n(GE)n(RE)n levels increased by stress are reduced by metformin. An interrupted *CASP8*-GGGAGA^{EXP} sequence variant is associated with increased AD risk (OR 2.2, $P = 3.1 \times 10^{-5}$). In transfected cells, a high-risk AD variant increases toxicity and poly(GR)n(GE)n(RE)n aggregates. These data support a model in which the *CASP8*-GGGAGA^{EXP} combined with stress increases AD risk.

Alzheimer's disease (AD), the most common form of dementia, is characterized by progressive cognitive decline and affects >10% of people older than 65 y of age (1). Rare segregating mutations in amyloid precursor protein, presenilin 1 and *2* (*PSEN1* and *PSEN2*) cause ~10% of early-onset AD (2), the common apolipoprotein ε4 allele (*APOE4*) is the strongest genetic risk factor for AD (3), and additional common and rare variants across the genome influence late onset AD risk (4). Hallmark pathology in AD brains includes extracellular β-amyloid (Aβ) plaques, intracellular phosphorylated Tau (p-Tau)-positive neurofibrillary tangles (NFTs), dystrophic neurites, reactive gliosis, and neurodegeneration (5, 6). However, the underlying mechanisms of most forms of AD are unclear and the genetics of AD are not fully understood.

Microsatellite repeats make up ~3% of the human genome and microsatellite expansion mutations at various loci cause >67 neurological diseases (7, 8). Moreover, an intronic G4C2 expansion mutation in *C9orf72* is the most frequent known genetic cause of both amyotrophic lateral sclerosis and frontotemporal dementia (C9 ALS/FTD) (9, 10). Molecular mechanisms of repeat expansion diseases include protein loss-of-function (LOF), protein gain-of-function (GOF), RNA GOF, and the accumulation of toxic polymeric proteins that can be expressed by repeat-associated non-AUG (RAN) translation (11). RAN proteins can be expressed from microsatellite expansion mutations in all reading frames without requiring AUG or AUG-like close cognate initiation codons (12). RAN

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: lien.nguyen@ufl.edu or ranum@ufl.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2416885122/-/DCSupplemental.

Published February 12, 2025.

proteins have been reported in 11 different repeat expansion disorders, and in vitro and in vivo studies support their pathologic role in a growing number of diseases (11, 13–20). Interestingly, pathological markers for a number of microsatellite expansion diseases (e.g. Huntington's disease, ataxia, and myotonic dystrophy) overlap with AD including increased p-Tau and the accumulation of NFTs (21–24).

The abundance of repetitive elements in the human genome raises the possibility that repeat expansions contribute to diseases with unresolved genetic etiology including AD (25). However, identifying these cases through traditional methods is challenging as repeat expansion tracts are difficult to sequence and align back to the human genome, and even with long-read sequencing require a strong family history of disease. Computational tools to identify repeat expansions in short-read NGS data, such as STRetch and ExpansionHunter DeNovo, show promise but are only partially concordant with long-read sequencing data (26–32). New tools are needed to fully explore the contribution of repeat expansions to AD and other disorders.

Here, we test the hypothesis that unidentified repeat expansions produce polymeric RAN proteins that form aggregates in AD brains and increase disease risk. Additionally, we show protein-coding information in these aggregates can be used to design guide RNAs, which in turn can be used to enrich and identify disease-causing mutations directly from patient genomic DNA using CRISPR/deactivated Cas9. Using this approach, we identified an interrupted intronic GGGAGA repeat expansion in the caspase 8 gene (*CASP8*), which encodes a cysteine-aspartic acid protease. We show that a specific expansion variant of this intronic *CASP8* GGGAGA expansion increases AD risk.

Results

polyGR Protein Frequently Accumulate in AD Autopsy Brain Tissue. Because RAN proteins are found in the most common known genetic form of FTD (33, 34), we tested the hypothesis that similar RAN proteins may accumulate in AD autopsy brains. We performed immunohistochemistry (IHC) using a previously developed polyclonal rabbit antibody against the polyGR repeat motif (a-polyGR), one of the most toxic *C9orf72* ALS/FTD dipeptide RAN proteins (14, 18, 20, 35). IHC staining using formalin-fixed hippocampal sections showed a-polyGR positive aggregates in sporadic AD but not age-similar control brains free of AD pathological changes (Braak stage ≤ 2, Thal stage ≤ 2)

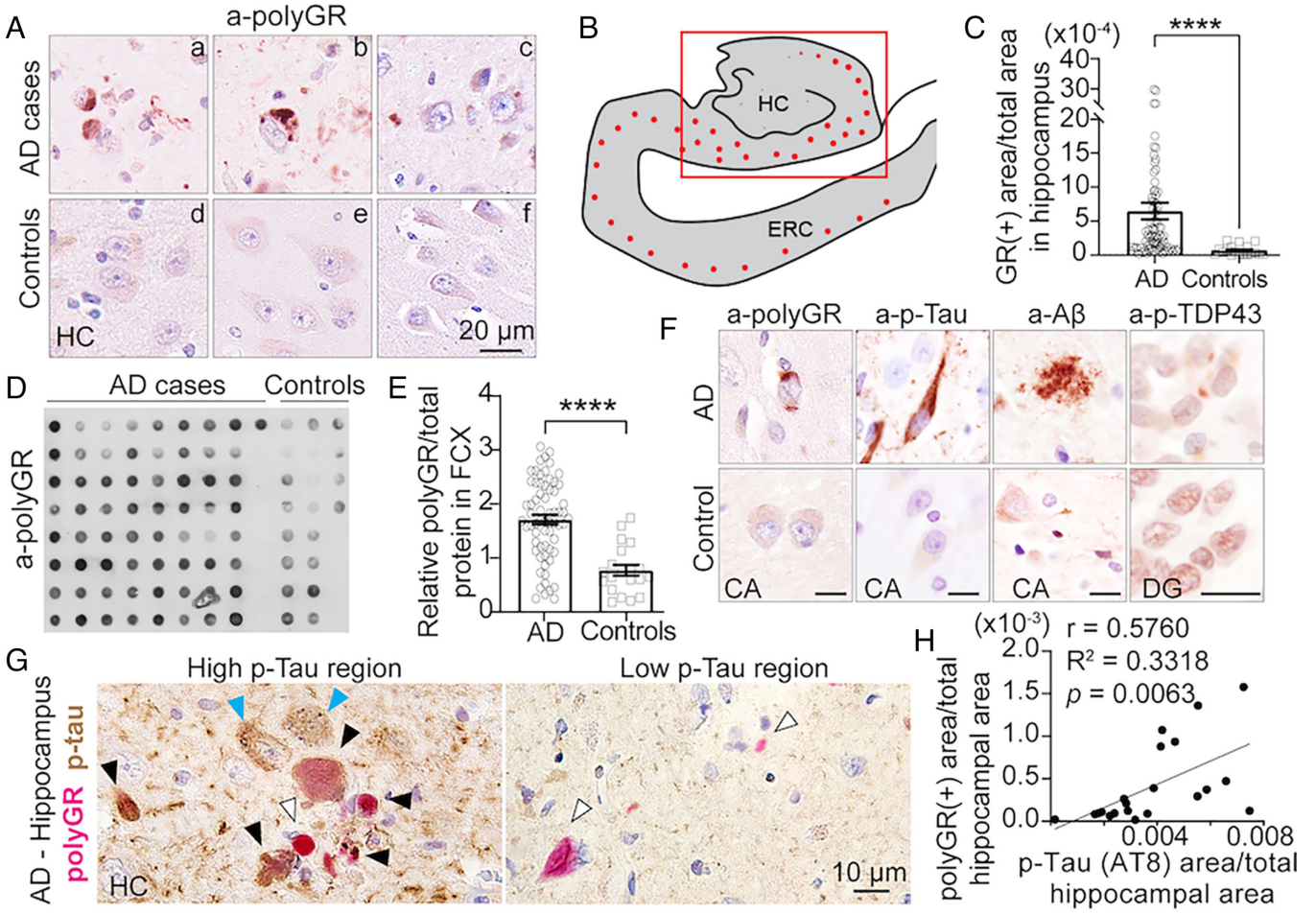

Fig. 1. polyGR aggregates accumulate in AD autopsy brains. (*A*) Example polyGR IHC staining (red) detected by a rabbit polyclonal a-polyGR antibody in hippocampal sections (HC) from AD and control cases. Each panel represents a different AD or control case. (*B*) Summary diagram showing location of polyGR aggregates (red dots) in AD HC sections; HC = hippocampus, ERC = entorhinal cortex. (*C*) Quantification of polyGR aggregates in HC region indicated by red square in (*B*) from AD (n = 80) and control (n = 18) cases. (*D*) Example dot blot of a-polyGR signal using a rat monoclonal a-polyGR antibody and protein extracts of the frozen frontal cortex (FCX) from AD (n = 65) and control (n = 20) cases. (*E*) Relative polyGR levels in FCX protein lysates from AD and controls by dot blot. (*F*) Example IHC staining showing distinct a-polyGR staining compared to other AD pathological hallmarks including p-Tau (S202 and T205), Aβ plaques, and p-TDP43 in Cornu Ammonis (CA) and dentate gyrus (DG) of the HC. Positive staining shown in red, (Scale bar, 20 μm.) (*G*) Examples of double IHC staining showing a-polyGR staining (pink) detected in brain regions with both high and low p-Tau (brown) subregions of the same AD brain section. Black arrows: cells with both polyGR and p-tau signal, open arrows: cells with polyGR staining, blue arrows: cells with p-Tau staining. (*H*) Plot of a-polyGR and p-Tau (S202 and T205) staining detected in sequential slides from 21 randomly selected AD cases shows a positive correlation. Data represent mean ± SEM. Two-tailed, unpaired *t* test. ****$P < 0.0001$.

(Fig. 1*A* and *SI Appendix*, Fig. S1*A*). To confirm antibody specificity, we show a-polyGR recognizes GR_{60} but not human tau in cells (*SI Appendix*, Fig. S1*B*). In a-polyGR(+) AD cases, a-polyGR staining showed frequent perinuclear aggregates in both neurons and glia in the hippocampal region, with the most prominent staining found in CA1 and subiculum areas (Fig. 1*B*). Quantification of a-polyGR IHC staining in hippocampal sections from 80 sporadic AD brains showed a-polyGR aggregates in 45 of 80 (56%) AD cases [age at death (AAD): 79.0 ± 11.9 (mean ± SD), 46.3% female, 53.7% male] compared to 18 age-similar controls [AAD: 78.0 ± 12.9 (mean ± SD), 61.9% female, 38.1% male] (95% CI [3.3×10^{-4} to 8.2×10^{-4}], *t*-statistic = 4.696, $P = 1.05 \times 10^{-5}$) (Fig. 1*C*). While minimal background staining in controls was detected in automated quantification (Fig. 1*C*), no similar a-polyGR aggregates were found in any of the 18 control samples.

To further explore the accumulation of polyGR-containing proteins in AD, we performed protein dot blots of frontal cortex lysates with available frozen tissue from a subset of the AD patient cohort (n = 65) and age-similar controls (n = 20, with 18 as described above plus two additional controls). Consistent with the results above, protein lysate blots probed with a rat monoclonal a-GR antibody (34, 36) showed increased levels of polyGR signal above the highest levels found in controls in 31 of 65 (48%) AD cases (relative intensity of 1.82 to 3.06) compared to age-similar controls (relative intensity 0.18 to 1.74) [95% CI (0.6659, 1.2078), *t* statistic = 6.938, $P = 5.81 \times 10^{-9}$] (Fig. 1 *D* and *E* and *SI Appendix*, Fig. S1*C*). In contrast, IHC staining using a-polyGR on hippocampal tissue from other known tauopathy-related diseases was mostly negative (*SI Appendix*, Fig. S2). We observed rare, mostly pin-like punctate aggregates in a few cases (1/9 LBD, 1/10 PSP, 3/10 Pick's), which were distinct from the more frequent, larger aggregates found in polyGR(+) AD cases.

Next, we compared the polyGR-containing aggregates found in AD cases with known pathological hallmarks of AD by IHC using serial hippocampal sections from the same AD cases. These studies show a-polyGR aggregates accumulate in distinct patterns compared to intracellular p-Tau tangles (AT8), extracellular Aβ plaques, or the much smaller p-TDP43 aggregates (e.g. Fig. 1*F*). Double IHC staining shows polyGR and p-Tau (AT8) staining is found in both overlapping (Fig. 1 *G*, *Left*) and distinct (Fig. 1 *G*, *Right*) subregions of the hippocampus with some cells positive for polyGR alone, p-Tau alone, or both polyGR and p-Tau (e.g. Fig. 1*G*). IHC staining of serial hippocampal sections shows a-polyGR aggregate staining is correlated with p-Tau accumulation and higher levels of polyGR aggregates are found in AD hippocampal regions with high levels of AT8 p-Tau (r = 0.5760, $R^2 = 0.332$, $P = 0.0063$) (Fig. 1*H*).

In summary, we show an unexpected type of a-polyGR(+) aggregates are found in ~50% of tested sporadic AD autopsy brains, and increased a-polyGR aggregate staining correlates with increased levels of p-Tau staining.

Development of dCas9READ to Identify PolyGR Encoding Repeat Expansions. The identification of a-polyGR aggregates in a large fraction of sporadic AD autopsy brains that are negative for *C9orf72* and SCA36 expansions suggested the presence of one or more unidentified GR-expressing repeat expansion mutations. To identify putative polyGR-encoding expansion mutations, we developed a repeat expansion enrichment method we call *CRISPR*/deactivated Cas9-based repeat enrichment and detection (dCas9READ) (Fig. 2*A*). dCas9READ works on the principle that putative repeat expansion mutations provide more binding sites for repeat-containing single guide RNAs (rsgRNA)/dCas9 complexes than shorter repeats and pull down of these complexes would allow the identification of repeat expansion DNA and corresponding unique flanking sequences. In proof-of-concept studies, dCas9READ pulldown with GGGGCC and CAGG rsgRNAs resulted in a 4- to 6-fold increase by quantitative PCR of *C9orf72* and *CNBP* flanking sequences in C9 ALS/FTD and myotonic dystrophy type 2 (DM2) patient samples, respectively, compared to controls (*SI Appendix*, Fig. S3 *A–D*). Similarly, Illumina sequencing identified the *C9orf72* G4C2 and *CNBP* CCTG expansion mutations and flanking regions as the most enriched sequences from C9 ALS/FTD and DM2 patient samples, respectively, after dCas9READ pulldown (*SI Appendix*, Table S1). As expected, no enrichment of the *C9orf72* G4C2 locus was found without the G4C2 rsgRNA. These results demonstrate that the repeat containing guide RNA is required for pulldown specificity (*SI Appendix*, Fig. S3*E*). Additionally, dCas9READ experiments performed using a mixture of 8 or all 24 possible rsgRNAs that would target polyGR encoding hexanucleotide repeats (2) successfully enriched the *C9orf72* G4C2 repeat locus (*SI Appendix*, Fig. S3*F* and Table S2). In summary, we show dCas9READ successfully enriched the *C9orf72* G4C2 and *CNBP* CCTG expansion mutations and corresponding unique flanking sequences directly from the genomic DNA of C9 ALS/FTD and DM2 patients. Additionally, our results highlight the utility of this approach to enrich and identify candidate GR-encoding repeat expansion mutations from patient DNA samples using pools of all possible GR-encoding rsgRNAs.

Identification of an Intronic GR-Encoding GGGAGA Repeat Expansion. Next, we used dCas9READ and a mixture of all possible GR-encoding sgRNAs (*SI Appendix*, Table S2) to pull down putative GR-encoding repeat expansions from genomic DNA isolated from five a-polyGR(+) AD cases and as well as an a-polyGR(−) AD and a non-AD sample as controls (Fig. 2*B*). Fragment analysis of enriched restriction enzyme digested genomic DNA showed a peak molecular weight of ~6 to 8 kb (*SI Appendix*, Fig. S4*A*). We performed short-read Illumina and PacBio no-amplification long-read sequencing on enriched DNA samples as described in *SI Appendix*, Fig. S4*B*. Short-read sequencing analysis identified 2,024 putative GR repeat loci with ≥ twofold enrichment over flanking sequence background in one or more AD samples (*SI Appendix*, Fig. S4*C*). Of these, 58.2% were intergenic, 33.1% intronic, and 8.7% were located in coding exons, 5′ or 3′ UTRs. These candidate loci were further prioritized based on fold enrichment of the repeat in one or more AD cases compared to control (*SI Appendix*, Table S3). An analysis of enriched vs. nonenriched loci for individual AD cases identified 19 loci with >twofold enrichment in one or more a-polyGR(+) AD cases compared to control (*SI Appendix*, Table S3) with *CASP8* by far the most highly enriched locus (4.97× and 9.63× fold enrichment) in two of the six AD cases (Fig. 2*C* and *SI Appendix*, Fig. S4*D*). Long-read PacBio sequencing of this region identified an interrupted intronic GGGAGA•TCTCCC candidate GR-encoding repeat expansion in intron 8 of *CASP8* (*CASP8*-$GGGAGA^{EXP}$). This candidate repeat is located within a composite repetitive element containing a short interspersed nuclear element with a retroviral origin (SINE-R), variable number tandem repeat (VNTR), and Alu-like sequences (SVA) type E referred to as an SVA-E element (Fig. 2*D*). The *CASP8* GGGAGA repeat in the reference genome (hg19 and hg38) contains seven GGGAGA units interspersed with interruptions (Fig. 2*D*). Long-read PacBio no-amplification sequencing reads detected *CASP8*-SVA expansions ranging in length from 44 to 64 GGGAGA repeat units in AD cases and controls (*SI Appendix*, Fig. S4*E*). These repeats are variably

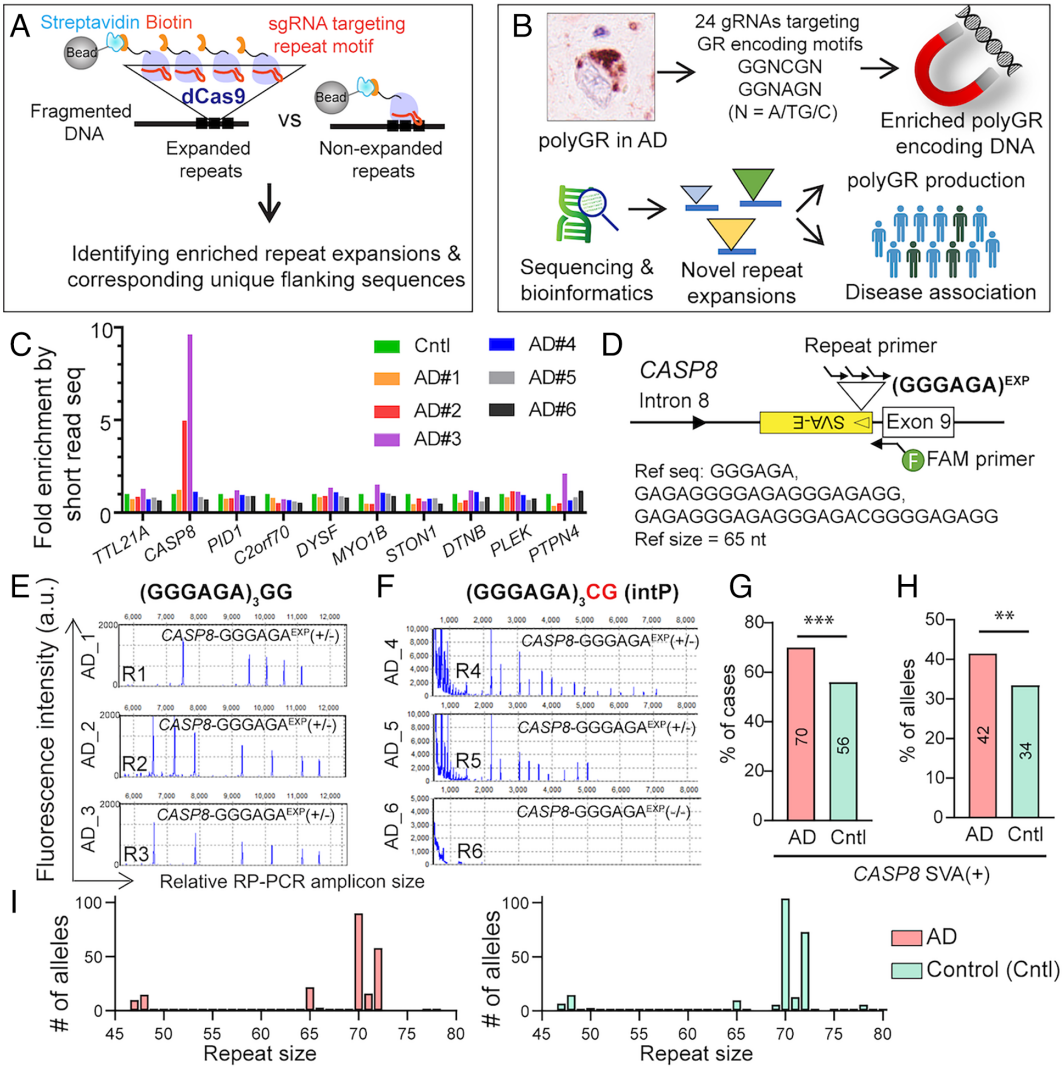

Fig. 2. Identification of interrupted GGGAGA repeat expansion in *CASP8* (*CASP8*-GGGAGA^{EXP}) using CRISPR deactivated Cas9-based repeat enrichment and detection (dCas9READ). (*A*) Schematic diagram showing dCas9READ strategy for enrichment of repeat expansion mutations encoding specific RAN protein motifs. (*B*) Experimental flow using dCas9READ to identify novel polyGR-encoding repeat expansions and genetic association studies to test the association of candidate mutations with AD. (*C*) Example data showing fold enrichment of ten loci after dCas9READ. Genomic DNA from five polyGR(+) AD cases, a polyGR(-) control (Cntl), a polyGR(−) AD (AD#4) was used. *CASP8* was the top hit in two polyGR(+) AD cases. (*D*) Diagram showing insertion of the *CASP8*-GGGAGA^{EXP} within an SVA-E retrotransposon element which is inserted in the opposite orientation with reference repeat sequence and repeat-prime PCR (RP-PCR) primers used to characterize the repeat expansion. (*E* and *F*) RP-PCR data showing five different repeat expansion patterns of *CASP8*-GGGAGA^{EXP} in 5 AD cases using $(GGGAGA)_3GG$ (*E*) or interrupted primer intP $(GGGAGA)_3CG$ (*F*). (*G*) Percentage of individuals carrying *CASP8*-SVA in initial AD and control cohorts. (*H*) Frequency of alleles containing *CASP8*-SVA insertion in AD (n = 266) and controls (n =400). (*I*) Distribution of estimated GGGAGA repeat lengths measured by LR-PCR in initial AD and control cohorts. (*G* and *H*) Chi-square test, $^{**}P < 0.01$, $^{***}P < 0.001$.

interrupted by multiple motifs (C, G, CG, CGG, GGG, GGGA & CGGG, and/or combinations of these motifs), which makes the overall repeat tract lengths longer.

CASP8 encodes a member of the cysteine-aspartic acid protease (caspase family). Rare protein-coding variants of *CASP8* [K148R, 1/600 in AD and 1/1,500 in controls; I298V, 1/300 in AD and 1/600 in controls] were previously associated with an increased risk of AD (37). These variants were not detected by PCR using primers listed in *SI Appendix*, Table S4 and Sanger sequencing of DNA from 24 *CASP8*-GGGAGA^{EXP}(+) cases. Given the frequent polyGR pathology in AD cases, the highest fold enrichment of the *CASP8* SVA-E sequence by dCas9READ in two out of six enriched AD cases, and the presence of an expanded GR encoding GGGAGA repeat in *CASP8* SVA-E, additional studies were focused on characterizing the potential role of the *CASP8*-GGGAGA^{EXP} in AD.

Repeat primed PCR (RP-PCR) using a FAM primer (Fig. 2*D*) targeting unique sequence downstream of the *CASP8*-GGGAGA^{EXP} with $(GGGAGA)_3GG$ primer showed signal (Fig. 2*E* and *SI Appendix*, Fig. S5*A*), but not the expected even saw-tooth pattern of previously published microsatellite expansion mutations with pure repeat tracts (e.g. *C9orf72* (G4C2)^{EXP}) (38). Because our long-read sequencing data and the reference genome indicated the *CASP8*-SVA-E repeat expansion contains GGGAGA repeats with sequence interruptions, we modified the RP-PCR protocol to use an interrupted repeat primer (intP) $(GGGAGA)_3CG$ that allowed binding to a commonly found interrupted sequence within the repeat tract. RP-PCR analysis using intP showed repeat expansions with a variety of distinct interruption patterns (e.g. Fig. 2*F*). It was previously shown that the *CASP8* SVA-E and approximately 300 bp of unique upstream sequence from *CASP8* intron 8 was transduced to chromosome 19 (39). To ensure we were analyzing the *CASP8* SVA insertion on chromosome 2, we used primers that amplify the *CASP8* SVA-E insertion on chromosome 2 but not the transduced SVA-E element located on chromosome 19 (*SI Appendix*, Fig. S5*B*).

SVA-E elements are one of the most recent types of transposable elements to be inserted into the human genome. Consistent with a recent insertion event, analysis of a tagging SNP [rs1035142(T)] in linkage disequilibrium with the *CASP8* SVA-E insertion (r^2 = 0.999) shows 53 to 65% of alleles tested in the 1,000Genomes, TWINSUK, and *All of Us* datasets do not have the *CASP8* SVA-E insertion. The sum of the frequencies of alleles in individuals with 0, 1, or 2 copies of the *CASP8* SVA-E insertion is equal to 1 (*SI Appendix*, Fig. S5*C*). Long-range PCR (LR-PCR) across the GGGAGA repeat expansion (*SI Appendix*, Fig. S5*D*) or SVA element (*SI Appendix*, Fig. S5*E*) showed the *CASP8*-GGGAGA^{EXP} is overrepresented in AD compared to controls in our initial cohort with approximately 70% of the tested AD (186/266) and 56% of control cases (n = 224/400) positive for one or both alleles (Fig. 2*G*) (P = 0.0029). Similarly, the overall allele frequency of the *CASP8*-GGGAGA^{EXP} was higher in AD cases (42%,

n = 221/532) compared to age-similar controls (34%, 268/800) (P = 0.0003) (Fig. 2*H*).

To examine the effect of the *CASP8* SVA insertion in an independent unbiased dataset we evaluated two SNPs in linkage disequilibrium with the SVA insertion using a dataset with 344 individuals with AD and 138,365 non-AD individuals from the NIH *All of Us* dataset (40). While first SNP, which tags all of the *CASP8* SVA insertions (rs1035142) is not associated with an increased AD risk [OR of 1.13 (SE = 0.080, beta = 0.12, P = 0.129)], the second SNP (rs700635) which is found on subset of *CASP8* SVA (r2 = ~0.5) is associated with increased risk of AD with an OR of 1.21 (SE = 0.088, beta = 0.19, P = 0.032) after adjusting for age, sex, and the first three principal components.

LR-PCR and RP-PCR show similar repeat length (*SI Appendix*, Fig. S5*F*) and interruption patterns (*SI Appendix*, Fig. S5*G*) in genomic DNA isolated from the blood and frontal cortex from two AD cases suggesting minimal somatic repeat instability. Repeat lengths measured by LR-PCR showed *CASP8*-GGGAGA repeats have ~47 to 76 GGGAGA repeats in AD and control cases in our initial cohort (Fig. 2*I*). qRT-PCR results showed a nonsignificant trend toward increased *CASP8* transcript levels in end-stage frontal cortex autopsy tissue from a small cohort of *CASP8*-GGGAGA^{EXP}(+) versus *CASP8*-GGGAGA^{EXP}(–) cases (*SI Appendix*, Fig. S5 *H–L*).

In summary, we used dCas9READ to enrich for GR encoding repeat expansions using genomic DNA from polyGR(+) AD brains and identified an intronic GGGAGA^{EXP} in *CASP8*-SVA as a candidate noncoding repeat expansion that may produce polyGR-containing proteins and contribute to the polyGR(+) aggregates found in AD autopsy brains.

Polymeric Proteins Expressed from the *CASP8*-GGGAGA^{EXP} Accumulate in AD Autopsy Brains. Analysis of the upstream flanking sequence of the *CASP8*-SVA expansion predicts sense RAN proteins in two frames and in the third reading frame an upstream ATG could potentially be used (Fig. 3*A*). Additionally, sequence interruptions within the repeat tract are predicted to shift the reading frame resulting in the expression of chimeric glycine-arginine (GR), arginine-glutamic acid (RE), and glycine-glutamic acid (GE) repeat [(GR)n(RE)n(GE)n] proteins with unique C-terminal regions for each reading frame (Fig. 3*A*). Similar chimeric proteins leucine-proline (LP), leucine-serine (LS), and serine-proline (SP) or [(LP)n(LS)n(SP)n] may also be expressed in the antisense direction. To directly test whether sense polymeric proteins expressed from the *CASP8*-GGGAGA^{EXP} locus accumulate in a-polyGR(+) AD autopsy brains, we generated antibodies to the C-terminal epitopes for *CASP8* sense frame 1 (Sf1) and sense frame 3 (Sf3) (CT-Sf1 and CT-Sf3 respectively) (Fig. 3*A*). Antibody validation experiments show *CASP8* Sf1 (a-CT-Sf1$_S$) and Sf3 (a-CT-Sf3$_S$) C-terminal antibodies colocalize with 5' FLAG epitope-tagged recombinant proteins expressing the corresponding unique C-terminal regions from frames 1 and 3 (*SI Appendix*, Fig. S6). No similar signal was detected in cells

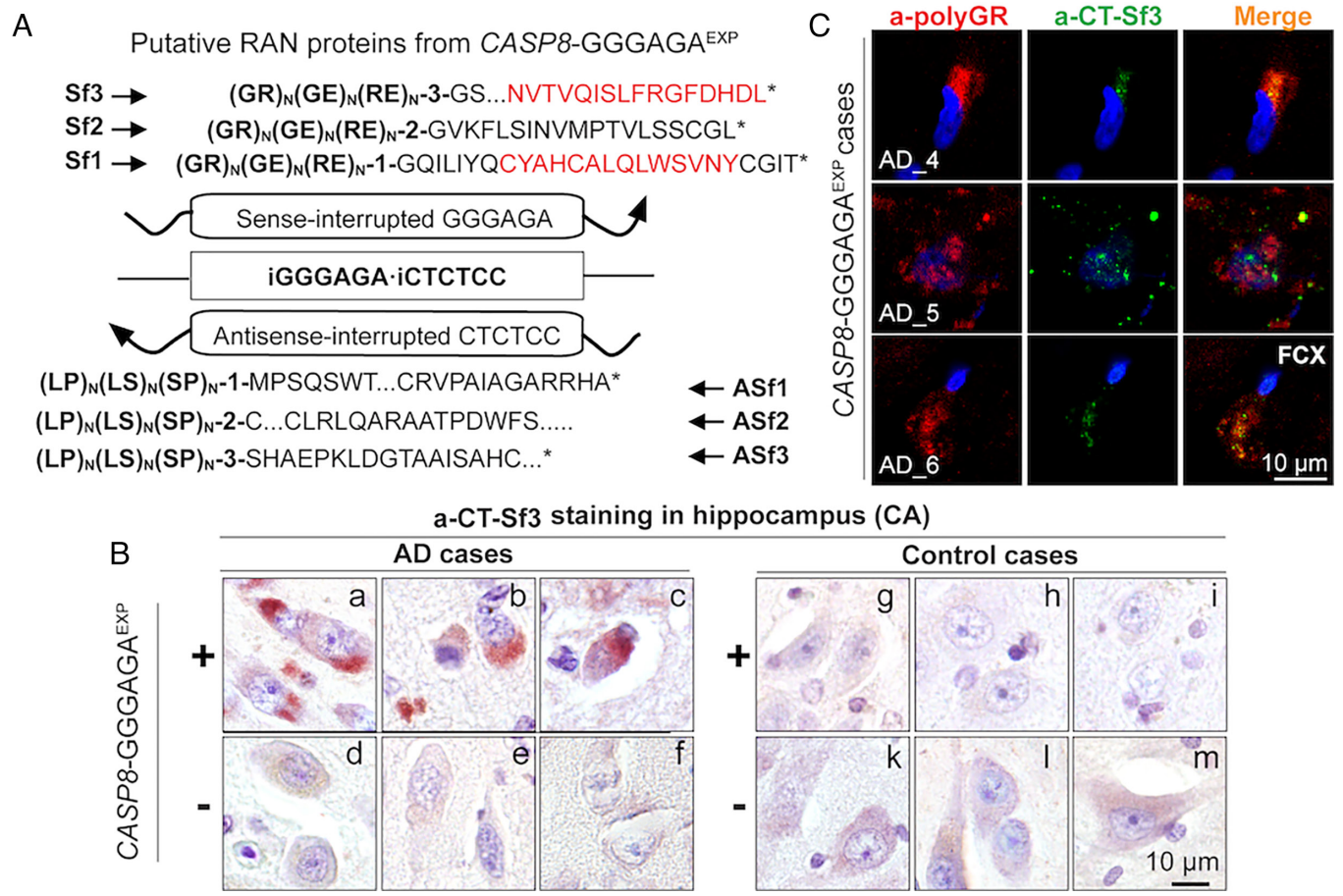

Fig. 3. *CASP8*-GGGAGA^{EXP} produces polyGR(+) aggregates in *CASP8*-GGGAGA^{EXP}(+) AD brains. (*A*) Diagram showing putative expansion proteins translated from sense and antisense *CASP8*-GGGAGA^{EXP} transcripts. Amino acid sequences highlighted in red were used to generate frame-specific C-terminal (CT) antibodies. S, sense; AS, antisense; f1–3, reading frames 1 to 3; * stop codon. Analysis of the upstream flanking sequence of *CASP8*-GGGAGA^{EXP} showed an upstream ATG that could be used in one of the three frames; however, this ATG-containing frame shifts reading frame depending on the highly polymorphic interruptions within the repeat tract. (*B*) IHC detection of *CASP8* RAN Sf3 protein aggregate staining (red) in the hippocampus of *CASP8*-GGGAGA^{EXP}(+) AD patients detected with a-CT-Sf3 antibody. Each photo panel represents a different AD or control case. (*C*) Double IF showing colocalization of a-polyGR (red) and a-CT-Sf3 (green) staining in the frontal cortex (FCX) of *CASP8*-GGGAGA^{EXP}(+) AD patients. Each row represents a different AD case.

transfected with control plasmids or in cells overexpressing FLAG-tagged CT-Sf1 or CT-Sf3 incubated with preimmune serum (*SI Appendix*, Fig. S6). These data demonstrate that a-CT-Sf1 and a-CT-Sf3 antibodies recognize their respective *CASP8* CT-Sf1 and CT-Sf3 targets.

IHC experiments using these antibodies were performed on hippocampal and frontal cortex sections using AD and control brains. IHC using a-CT-Sf3 detected aggregate- or fibril-like staining in *CASP8*-GGGAGA^{EXP}(+) AD cases but not *CASP8*-GGGAGA^{EXP}(–) AD or *CASP8*-GGGAGA^{EXP}(±) unaffected controls (Fig. 3*B* and *SI Appendix*, Fig. S7). The a-CT-Sf3 staining was detected in 11 out of 11 tested *CASP8*-GGGAGA^{EXP}(+) AD cases while no signal was detected in 4 *CASP8*-GGGAGA^{EXP}(–) AD, 5 *CASP8*-GGGAGA^{EXP}(+) control, and 3 *CASP8*-GGGAGA^{EXP}(–) control [95% CI (10.5213, 31424.4787), $P = 0.0019$]. Statistical analysis was performed using the odds ratio calculation tool (41), to test whether a-CT-Sf3 signal is associated with *CASP8*-GGGAGA^{EXP}(+) AD cases compared to the group of *CASP8*-GGGAGA^{EXP}(–) AD cases and controls. In the hippocampus, a-CT-Sf3 staining varied among AD cases and was robustly and frequently found in some cases and less frequent in others (*SI Appendix*, Table S5). In the frontal cortex, a-CT-Sf3 and a-CT-Sf1 staining was found in gray and white matter regions from a subset of *CASP8*-GGGAGA^{EXP}(+) but not *CASP8*-GGGAGA^{EXP}(–) AD cases or *CASP8*-GGGAGA^{EXP}(±) controls (*SI Appendix*, Fig. S8 and Table S5). Double immunofluorescence (IF) experiments on FCX tissue showed a-polyGR (34, 36) staining partially colocalizes with a-CT-Sf3 staining (Fig. 3*C* and *SI Appendix*, Fig. S9). Taken together, these results demonstrate that GR-containing polymeric proteins are expressed from at least two reading frames across the *CASP8*-GGGAGA^{EXP} and that *CASP8* expansion proteins cause at least part of polyGR pathology found in *CASP8*-GGGAGA^{EXP}(+) AD cases.

Of the 36 polyGR(+) autopsy brains with DNA available for genotyping, 27 were *CASP8*-GGGAGA^{EXP}(+) and nine *CASP8*-GGGAGA^{EXP}(–). These data suggest additional GR-encoding AD expansion mutations remain to be identified. Additionally, of the 11 *CASP8*-GGGAGA^{EXP}(+) AD autopsy brains available for additional IHC experiments, all were positive for a-polyGR, a-CT-Sf1, or a-CT-Sf3 (*SI Appendix*, Table S5). Taken together, these data demonstrate that the *CASP8*-GGGAGA^{EXP} expresses polymeric proteins in *CASP8*-GGGAGA^{EXP}(+) AD brains and provides a molecular explanation for a large portion of the polyGR staining found in our AD autopsy brain set.

CASP8-GGGAGA^{EXP} Minigenes Express Poly(GR)n(GE)n(RE)n RAN Proteins in Cells. To test whether the *CASP8*-GGGAGA^{EXP} expresses poly(GR)n(GE)n(RE)n proteins, we performed a series of minigene experiments in transfected cells. These minigene constructs (C8-GGGAGA^{EXP}-3T) contain six upstream stop codons (2 in each reading frame), 100-bp of unique upstream sequence of the GGGAGA^{EXP} and one of several different interrupted *CASP8*-GGGAGA^{EXP} alleles identified by long-read PacBio sequencing (*SI Appendix*, Fig. S10), and 3′-epitope tags (FLAG, HA, or Myc) in each reading frame (Fig. 4*A*). These minigenes, contain interrupted GGGAGA repeats with 44, 63, or 65 repeats. Each of these clones are predicted to express hybrid poly(GR)n(GE)n(RE)n proteins. The endogenous sequence upstream of the repeat in the HA but not the other frames for each of these constructs contains a potential upstream ATG-initiation codon. Protein blots of HEK293T cells transfected with C8-GGGAGA^{EXP}-3T plasmids detected proteins in the FLAG and HA frames (Fig. 4*B*) and IF showed proteins are expressed in all three reading frames (Fig. 4*C*).

Because the interrupted *CASP8*-GGGAGA^{EXP} produces mixed poly[(GR)n(RE)n(GE)n] proteins with only short interspersed stretches of 3-7 consecutive GR repeats (*SI Appendix*, Fig. S11), we tested whether these hybrid repeat proteins can be detected by the a-polyGR antibody (14) when expressed in cells. Double IF showed colocalization of a-polyGR and a-FLAG staining in SH-SY5Y cells transfected with minigenes that express AUG-initiated FLAG-tagged *CASP8* [(GR)n(RE)n(GE)n] proteins (*SI Appendix*, Fig. S11). This result demonstrates that in addition to detecting pure polyGR tracts that are expressed in *C9orf72* (14), the a-polyGR antibody (14) also detects interrupted polyGR stretches found in the hybrid *CASP8* [(GR)n(RE)n(GE)n] proteins. These data further support that *CASP8* polymeric proteins contribute to the a-polyGR pathology found in *CASP8*-GGGAGA^{EXP}(+) AD brains.

Stress Increases _CASP8_ RAN Protein Levels and PolyGR and _CASP8_ polyGR-GE-RE Proteins Increase p-Tau in Cells. Given the lack of staining in *CASP8*-GGGAGA^{EXP}(+) age-matched controls without typical AD pathology (Braak stage ≤ 2, Thal score ≤ 2), we hypothesize that stress or environmental factors, which are known to increase RAN translation (16, 42–46), may increase the expression of *CASP8* poly[(GR)n(RE)n(GE)n] proteins. To test this hypothesis, we examined the effects of thapsigargin, which increases endoplasmic reticulum (ER) stress (47, 48), on *CASP8* poly[(GR)n(RE)n(GE)n] protein levels. Thapsigargin-treated cells transfected with C8-GGGAGA^{EXP}-3T show increased levels of poly[(GR)n(RE)n(GE)n] of ~2 and ~1.5 fold in the FLAG ($P < 0.0001$) and HA ($P = 0.0002$) frames, respectively (Fig. 4 *D* and *E* and *SI Appendix*, Fig. S12*A*). Additionally, we show that metformin, an FDA-approved type 2 diabetes drug that was recently shown to decrease RAN protein levels expressed from other repeat expansions (GGGGCC, CCTG, CAGG, and CAG) (16), reduced poly[(GR)n(RE)n(GE)n] levels in thapsigargin-treated cells in the FLAG ($P < 0.0001$) and HA ($P < 0.0001$) reading frames (Fig. 4 *D* and *E* and *SI Appendix*, Fig. S12*A*) without changing the C8-GGGAGA^{EXP}-3T transcript levels (*SI Appendix*, Fig. S12*B*).

Because higher polyGR levels were detected in AD cases with higher p-Tau (S202 and T205) (Fig. 1*H*), we tested whether expression of polyGR or poly[(GR)n(RE)n(GE)n] proteins alter p-Tau levels in cells. Double IF shows SH-SY5Y cells transfected with plasmids containing alternative non-hairpin forming GR codons (3xFLAG-AltGR$_{60}$) plasmid (*SI Appendix*, Fig. S1*B*), have significantly higher p-Tau signal (AT8) compared to cells transfected with empty vector ($P < 0.0001$) (Fig. 4 *F* and *G*). Similarly, we detected increased p-Tau levels in poly[(GR)n(RE)n(GE)n](+) cells as C8-GGGAGA^{EXP}-3T was transfected in SH-SY5Y cells (Fig. 4*H*).

In summary, ER stress induced by thapsigargin increases *CASP8* poly[(GR)n(RE)n(GE)n] RAN protein levels, suggesting stress conditions favor the accumulation of *CASP8* RAN proteins. Overexpression of polyGR or *CASP8*-GGGAGA^{EXP}-containing constructs leads to increased levels of p-Tau. Additionally, our data highlight the potential beneficial effects of metformin, which decreased poly[(GR)n(RE)n(GE)n] protein levels in thapsigargin-treated cells.

A _CASP8_-GGGAGA^{EXP} Variant (_CASP8_-GGGAGA-AD-R1) Increases AD Risk. Because the *CASP8*-GGGAGA^{EXP} produces polyGR-containing polymeric poly[(GR)n(RE)n(GE)n] proteins in transfected cells and AD autopsy brains, we asked whether this repeat expansion is genetically associated with AD. Interestingly, we noticed a specific repeat configuration R1 (Fig. 2*E*) was frequently found in AD cases. This variant group, which contains the R1 repeat interruption pattern is distinct from other variants (e.g. R2-R5) (Fig. 2*E* and *SI Appendix*, Fig. S5*A*). Pacbio long-read amplicon sequencing of individuals with AD and the R1 variant identified an extra 48 to 50 bp repeat motif in

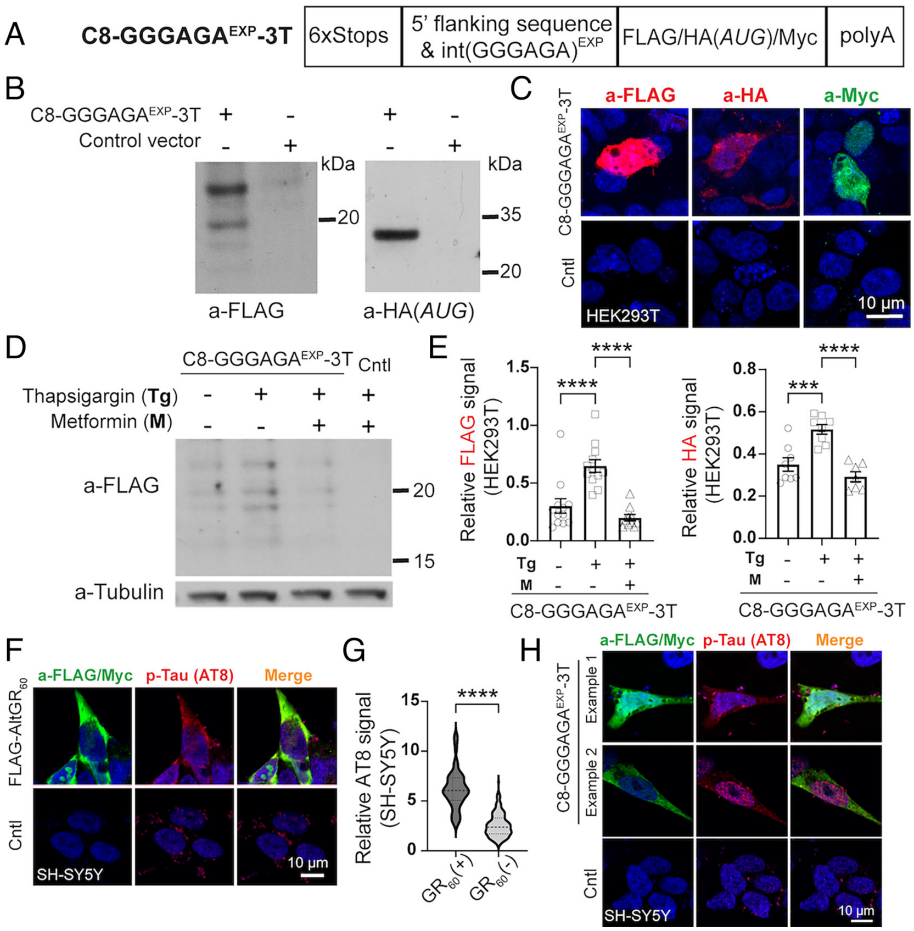

Fig. 4. *CASP8*-GGGAGA^{EXP} produces polyGR(+) proteins and increases p-Tau in transfected cells. (*A*) Diagram showing C8-GGGAGA^{EXP}-3T constructs containing 6xStop codons (two in each reading frame), a 100-bp of *CASP8* flanking sequence upstream of the repeat, interrupted GGGAGA repeat expansions, and triple epitope tags in each of the three reading frames. (*B*) Protein plot showing C8-GGGAGA^{EXP}-3T expresses mutant proteins in FLAG and HA frames in transfected HEK293T cells but not empty vector controls. (*C*) Mutant proteins in all three reading frames (FLAG, HA, and Myc) were detected by IF in HEK293T cells transfected with C8-GGGAGA^{EXP}-3T but not empty vector plasmids. (*D*) Protein blotting showing effects of Thapsigargin (Tg) and metformin (M) on the levels of *CASP8* RAN proteins expressed in the FLAG frame. (*E*) Quantification of *CASP8* RAN proteins expressed from the FLAG and HA frames with Tg or Tg + metformin treatment (n = 7 to 12/condition). (*F*) IF images showing increased p-Tau levels in SH-SH5Y cells expressing FLAG-tagged-$(GR)_{60}$ protein encoded using alternative codon sequences. (*G*) Quantification of p-Tau levels in polyGR(+) (n = 46) and polyGR(−) SH-SY5Y cells (n = 287), Cntl: empty vector control. (*H*) IF images showing increased p-Tau levels in SH-SY5Y cells transfected with C8-GGGAGA^{EXP}-3T plasmids. Data represent mean ± SEM. One-way ANOVA Holm–Sidak's multiple comparisons test (*E*) and unpaired two-tailed *t* test (G). ***$P < 0.001$, ****$P < 0.0001$.

the VNTR region (long VNTR), 3 to 4 units of [$(GGGAGA)_4CG$] within the repeat tract, and 5 to 7 units of (GGGAGA) at the 3′ end (Fig. 5 *A*–*C*). We refer to these sequence subvariants as *CASP8*-GGGAGA-AD-R1 (Fig. 5*B*). The long VNTR is found both in the *CASP8*-GGGAGA-AD-R1 subvariants and a common variant with a shorter hexamer region (Fig. 5*C*). Next, we used LR-PCR and RP-PCR that amplify the hexanucleotide repeat region to test whether the *CASP8*-GGGAGA-AD-R1 subvariants are associated with AD using genomic DNA samples from three independent AD patient cohorts along with age-similar cognitively unaffected White non-Hispanic controls. Cohort 1 was collected as part of the National Centralized Repository for AD and Related Dementias (NCRAD). The odds ratio (OR) of finding *CASP8*-GGGAGA-AD-R1 variants in AD vs. controls in Cohort 1 (n = 734 AD and n = 748 controls) was 1.8 (95% CI [1.1248, 2.8394], z statistic 2.458, $P = 0.0140$) (Fig. 5*D* and Table 1). Cohort 2 contained samples from Johns Hopkins University, 1Florida ADRC and Coriell. For cohort 2 (n = 278 AD, n = 449 controls) the OR of carrying a *CASP8*-GGGAGA-AD-R1 variant in AD vs. controls was 3.6 (95% CI [1.8284, 7.1186], z statistic 3.700, $P = 0.0002$) (Fig. 5*D* and Table 1). A third independent replication cohort, Cohort 3 from the Massachusetts General Hospital ADRC showed a similar trend with 10/162 AD cases (6.2%) vs 1/42 controls (2.4%) positive for the *CASP8*-GGGAGA-AD-R1 variant. For the combined cohort (n(AD) = 1,174, n(control) = 1,239), the OR of the *CASP8*-GGGAGA-AD-R1 variant in AD vs controls was 2.2 [95% CI (1.5185, 3.1896), z statistic 4.166, $P = 3.10\times10^{-5}$] (Fig. 5 and Table 1). In the combined cohorts, repeat lengths of the *CASP8*-GGGAGA-AD-R1 variant ranged in size from ~57 to 68 repeats.

It is important to highlight that the SVA insertions are hominid specific and that SVA-E subtype elements are human specific and evolutionarily recent (49). The *CASP8* insertion event is found on ~54% of human alleles and this insertion has continued to mutate at a higher rate than the flanking sequences. The proposed model shown in Fig. 5*E* illustrates the order in which the various *CASP8* SVA-E alleles flanking SNPs arose. The rs1035142 SNP, which is in complete LD with the *CASP8*-SVA-E insertion arose earlier than the rs700635 SNP which is evolutionarily closer to the R1 variant. Although the rs700635 SNP does not exclusively tag the R1 alleles it is a better tag for the longer VNTR, which is one of the molecular changes found on both the R1 allele and a subset of other more recent *CASP8* SVA alleles.

In addition to the association of the *CASP8*-GGGAGA-AD-R1 allele with AD in the 65+ population, in the 87+ age range we found *CASP8* SVA insertions are overrepresented in AD (235/341 – 68%) compared to controls (118/192 – 61%) [OR 1.39, 95% CI (0.8448 to 1.4884), $P = 0.08$]. While these data are not significant with this relatively small 87+ cohort, the trend suggests the need for further studies of the risk of *CASP8*-GGGAGA-AD-R1 and other *CASP8* SVA insertion alleles in AD.

Because *APOE4* is a strong AD risk factor, we tested whether the increased AD risk associated with the R1 variant is influenced by APOE alleles (*SI Appendix*, Table S6). No significant increases in R1 association with AD were detected in R1(+) AD cases with one or more ApoE4 alleles. However, in a subcohort of *APOE4* negative AD cases (n = 275) and controls (n = 550) from NCRAD, the association of the *CASP8*-GGGAGA-AD-R1 variant with AD was marginally increased with an OR of 2.5 [95% CI (1.3927, 4.5092), z statistic 3.065, $P = 0.0022$] (*SI Appendix*, Table S7). Taken together, these data suggest that the *CASP8*-GGGAGA-AD-R1 association with AD is independent of *APOE4*.

In summary, our results show that the *CASP8*-GGGAGA-AD-R1 repeat expansion variant is an AD risk factor.

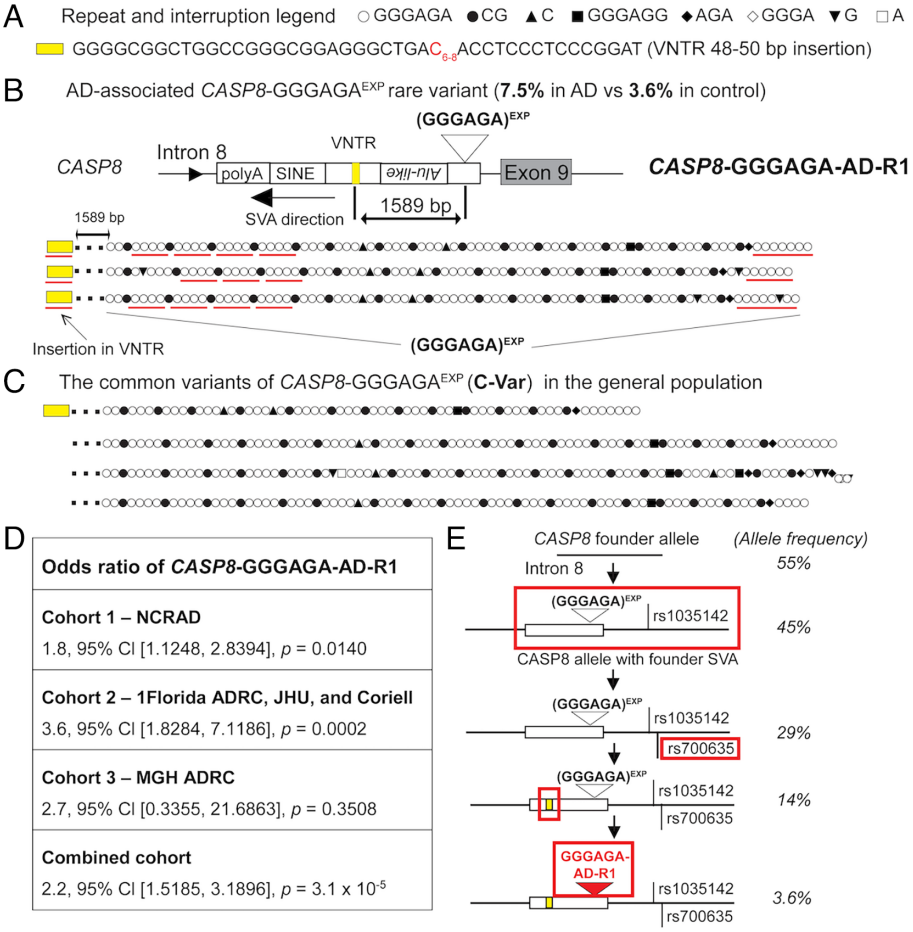

Fig. 5. *CASP8*-GGGAGA^{EXP} sequence variant increases AD risk. (*A*) Repeat legend showing repeat motifs and interruptions detected in *CASP8* SVA-E GGGAGA^{EXP} alleles. (*B*) Repeat configurations of AD-associated *CASP8* SVA-E GGGAGA^{EXP} variants (*CASP8* GGGAGA-AD-R1) which contains a longer VNTR region with a 48 to 50 bp insertion (yellow box) and interrupted GGGAGA repeat expansion including 3 to 4 units of [(GGGAGA)$_4$CG] within the repeat, and 5 to 7 units of (GGGAGA) at the 3′ end. The red lines highlight unique features in the *CASP8*-GGGAGA-AD-R1 variant compared to other repeat variants in the *CASP8* locus. (*C*) Repeat configurations of other *CASP8* GGGAGA^{EXP} variants found in the general population. (*D*) A summary of the association of *CASP8*-GGGAGA-AD-R1 variants with AD in three independent AD and control cohorts. (*E*) Proposed model showing the various CASP8 alleles in the population and the order in which the *CASP8* SVA-E insertion and tagging SNPs (rs1035142 and rs700635), the longer VNTR, and the interrupted hexanucleotide stretch on the AD-R1 alleles arose.

AD-Associated *CASP8*-GGGAGA-AD-R1 Shows Increased RAN Proteins, RNA Inclusions, and Cell Toxicity. To better understand the differences in expression of the *CASP8*-GGGAGA-AD-R1 and other *CASP8*-GGGAGA^{EXP} variants, we generated plasmids containing 6xSTOP codons (two in each reading frame) followed by GGGAGA^{EXP} sequences cloned from patients. *CASP8*-GGGAGA-AD-R1 (p-AD-R1 plasmids) or another *CASP8*-GGGAGA^{EXP} variant (p-C-Var plasmids) were transfected into HEK293T cells (Fig. 6*A*). We further characterized the *CASP8*-GGGAGA^{EXP} p-AD-R1 and p-C-Var variants using RP-PCR and LR-PCR (*SI Appendix*, Figs. S13 and S14). IF experiments show a-polyGR levels were 4.2X higher in cells transfected with high-risk p-AD-R1 repeats compared to the other *CASP8*-GGGAGA^{EXP} variant allele (p-C-Var) ($P = 0.0045$) (Fig. 6*B* and *SI Appendix*, Fig. S15). Plasmid

Table 1. Association of *CASP8*-GGGAGA-AD-R1 with AD in three independent cohorts

	CASP8-GGGAGA-AD-R1	All other variants*	Total cases	Odds ratio	95% CI	z statistic	*P* value
Cohort 1 – NCRAD							
AD	51 (6.9%)	683	734	1.8	1.1248, 2.8394	2.458	0.0140
Control	30 (4.0%)	718	748				
Cohort 2 – 1Florida ADRC, JHU, and Coriell							
AD	27 (9.4%)	251	278	3.6	1.8284, 7.1186	3.700	0.0002
Control	13 (2.8%)	436	449				
Cohort 3 – MGH ADRC							
AD	10 (6.2%)	152	162	2.7	0.3355, 21.6863	0.933	0.3508
Control	1 (2.4%)	41	42				
Total AD and control cohorts							
AD	88 (7.5%)	1,086	1,174	2.2	1.5185, 3.1896	4.166	3.1×10^{-5}
Control	44 (3.6%)	1,195	1,239				

Cohort 1 – NCRAD: 734 AD samples (51% female, 49% male, non-Hispanic, White, age at death (AAD) [(min, max), median ± SD] = [(65, 90+), 81.0 ± 5.1 y], and 748 cognitively healthy controls (50% female, 50% male, non-Hispanic, White, AAD = [(65, 90+), 78.0 ± 5.2 y]. >95% of genomic DNA samples are from blood. Cohort 2 – 1Florida ADRC, JHU, and the Coriell: 278 autopsy-confirmed cases from UF and JHU ADRCs (58% females and 42% males, non-Hispanic, White), genomic DNA samples are extracted from AD brain tissue, AAD = [(65, 102), 83.0 ± 7.4 y], and 449 cognitively healthy Coriell controls (48% females and 52% males, race: non-Hispanic, White), genomic DNA are from blood samples, Age at sampling (AAS) [(min, max), median ± SD] = [(65, 95), 73 ± 5.9 y). Cohort 3 – MGH ADRC cohort: 162 AD samples (53% female, 47% male, non-Hispanic, White, AAD = [(65, 90+), 81 ± 6.6 y] and 42 Control samples (49% female, 51% male, non-Hispanic, White AAD = [(73, 90+), 77.5 ± 5.8 y]. Genomic DNA samples were extracted from brain tissue.
*All other variants include non-R1 CASP8-GGGAGAEXP variants and cases without CASP8 SVA-E insertions found at the CASP8 locus.

RNA levels were comparable between the experimental groups (*SI Appendix*, Fig. S16). To understand the effects of the repeat interruptions on RNA foci/condensate formation, fluorescence in situ hybridization (FISH) using a $(TCTCCC)_3$ probe was performed. In p-AD-R1 transfected cells, a higher number of large and intense rGGGAGA^{EXP} RNA foci or condensates were observed (Fig. 6*C* and *SI Appendix*, Fig. S17) compared to p-C-Var transfected cells ($P = 0.0085$). Additionally, the p-AD-R1 plasmid is more toxic compared to p-C-Var (Fig. 6*D*). Transfection of p-AD-R1 or p-C-Var plasmids increased cell death by ~20X ($P = 0.0032$) or 10X ($P = 0.046$) compared to empty vector controls, respectively (Fig. 6*D*). p-AD-R1 transfected T98 cells showed a ~2X increase in cell death compared to cells transfected with p-C-Var plasmids ($P = 0.029$).

Interestingly, examination of the *CASP8*-GGGAGA^{EXP} variants in the a-polyGR(+) and a-CT-Sf3 positive autopsy brains shown in *SI Appendix*, Table S5 were positive for *CASP8*-GGGAGA-AD-R1 (1/11) and other *CASP8*-GGGAGA^{EXP} variants (10/11). Additionally, 4/11 of these cases were homozygous for the *CASP8*-GGGAGA^{EXP}, with one of these cases carrying both a *CASP8*-GGGAGA-AD-R1 and another *CASP8*-GGGAGA^{EXP} variant allele. While our association data demonstrate that the *CASP8*-GGGAGA-AD-R1 variant increases AD risk (OR 2.2, $P = 3.10 \times 10^{-5}$), analysis of *CASP8*-GGGAGA^{EXP} autopsy tissue shows that both *CASP8*-GGGAGA-AD-R1 and other *CASP8*-GGGAGA^{EXP} alleles can express CASP8-locus-specific RAN proteins that accumulate in *CASP8*-GGGAGA^{EXP} positive AD brains. Stress increases RAN translation of *CASP8*-GGGAGA^{EXP} (Fig. 4 *D* and *E*) and other expansion mutations (16, 42, 50–52). These results combined with the lack of a-polyGR and a-Ct-Sf3 staining in *CASP8*-GGGAGA^{EXP} control brains suggests that *CASP8*-GGGAGA^{EXP} combined with stress or other genetic or environmental risk factors that trigger *CASP8* RAN translation contribute to the development and progression of the full disease.

Taken together, our results support a model in which the *CASP8*-GGGAGA-AD-R1 variant leads to higher AD-risk because it produces increased levels of RAN protein aggregates and RNA inclusions. RAN protein aggregates in turn, increase p-Tau. Additionally, stress, which is known to increase RAN protein levels (16, 42, 50–52), increases the expression and accumulation of poly[(GR)n(RE)n(GE)n] RAN aggregates from both the *CASP8*-GGGAGA-AD-R1 and common *CASP8*-GGGAGA^{EXP} variants which contribute to disease (Fig. 6*E*).

Discussion

We demonstrate that a-polyGR(+) aggregates are a frequent and unexpected pathological feature commonly found in sporadic AD cases but not age-similar control autopsy brains. Anti-polyGR(+) aggregates are distinct from Aβ, p-Tau, and TDP43 aggregates. Hippocampal regions with high levels of polyGR(+) aggregates also show increased p-Tau. We developed dCas9READ as a tool to enrich microsatellite repeat expansions directly from a-polyGR(+) AD patient brains, and used this method to identify an interrupted GGGAGA repeat expansion in intron 8 of *CASP8* (*CASP8*-GGGAGA^{EXP}). This interrupted GGGAGA^{EXP}, which is located within an SVA-E element oriented in the opposite direction to *CASP8* within intron 8, expresses hybrid poly[(GR)n(RE)n(GE)n] RAN proteins. IHC using a-polyGR and locus-specific C-terminal antibodies show that the hybrid poly(GR)n(GE)n(RE)n proteins expressed from the *CASP8*-GGGAGA^{EXP} form aggregates in *CASP8*-GGGAGA^{EXP}(+) AD brains. In cells, stress-induced *CASP8*-RAN protein levels are reduced by metformin, an FDA-approved type-2 diabetes drug that was recently shown to lower RAN-protein levels and increase motor neuron survival in *C9orf72* ALS/FTD BAC mice (16). Additionally, we found that a specific *CASP8* expansion variant *(CASP8*-GGGAGA-AD-R1) is associated with AD in White non-Hispanic 65+ cohorts compared to age-similar controls, OR 2.2 [95% CI (1.5185, 3.1896), $P = 3.10 \times 10^{-5}$]. Finally, we show that the *CASP8*-GGGAGA-AD-R1 variant produces higher levels of poly[(GR)n(RE)n(GE)n] RAN proteins and RNA inclusions and is more toxic to cells than a common *CASP8*-GGGAGA^{EXP} variant. In summary, we identify a novel intronic GGGAGA expansion in *CASP8* as a risk factor for AD and show that this poly[(GR)n(RE)n(GE)n] encoding repeat expresses RAN proteins that explain part of the polyGR(+) staining seen in AD autopsy brains. Additionally, we show that dCas9READ, a RAN pathology to gene identification strategy developed here, enables the unbiased identification of expansion mutations directly from genomic DNA of RAN-positive patient samples.

The accumulation of intracellular pTau tangles is a pathological hallmark of AD; however, the cause of this pathology in most sporadic AD cases is unclear. Our data show hippocampal regions with increased a-polyGR(+) staining also have increased p-Tau in AD autopsy brains. Additionally, overexpression of polyGR or *CASP8* poly[(GR)n(RE)n(GE)n] proteins increases p-Tau levels in cells. Although additional studies are needed to understand the distribution of polyGR(+) aggregates and p-Tau throughout *CASP8*-GGGAGA^{EXP}(+) AD brains, our data suggest the possibility that *CASP8* RAN proteins trigger p-Tau accumulation. It is also possible that p-Tau accumulation creates stress or other conditions that further favor the accumulation of poly[(GR)n(RE)n(GE)n] proteins. Either scenario could create a positive feedback loop that increases RAN and p-Tau pathology and worsens disease.

In the related neurodegenerative disorder *C9orf72*-linked ALS/FTD, polyGR aggregates have been reported to induce TDP-43 proteinopathy (53). Given 7/10 *CASP8*-GGGAGA-AD-R1(+) AD brains from MGH also show TDP43 pathology, future studies aimed at understanding the contributions that *CASP8* poly[(GR)n(RE)n(GE)n] RAN proteins and *CASP8* expansion RNAs have on p-TDP43 and additional aspects of AD pathology (e.g. Aβ accumulation) and vice versa will provide further insight into the molecular mechanisms of AD. It is possible that *CASP8* poly[(GR)n(RE)n(GE)n] proteins have shared toxic properties with *C9orf72* arginine-containing RAN proteins (20, 54–58) although the hybrid nature of the repeat motif is also likely to confer distinct cellular effects.

In contrast to the a-polyGR(+) and *CASP8*-GGGAGA^{EXP}(+) AD cases, all five *CASP8* GGGAGA^{EXP}(+) control brains were negative for a-polyGR and *CASP8* locus-specific C-terminal antibody staining. These data suggest the possibility that both the *CASP8*-GGGAGA^{EXP} and one or more additional factors are required to trigger *CASP8* RAN protein production and accumulation in AD brains. These results combined with the observation that *CASP8* poly[(GR)n(RE)n(GE)n] RAN protein levels are upregulated under ER stress in cells suggest stress conditions may increase susceptibility to AD in *CASP8*-GGGAGA^{EXP}(+) carriers by increasing *CASP8* RAN protein levels. This hypothesis is consistent with other studies showing a role for oxidative stress in AD (59–61).

Among the polyGR-positive sAD autopsy brains that we were able to genotype, 27/36 were positive for the *CASP8*-GGGAGA^{EXP}. Two different *CASP8*-locus-specific C-terminal antibodies showed similar staining patterns in *CASP8*-GGGAGA^{EXP}(+) AD brains. These data combined with the colocalization of a-polyGR staining with staining from the *CASP8*-locus-specific C-terminal antibody a-CT-Sf3, demonstrates that the *CASP8*-GGGAGA^{EXP} causes a substantial portion of the polyGR pathology detected in a-polyGR(+) AD autopsy brains. Interestingly, 9/36 a-polyGR(+) AD autopsy brains were negative for the *CASP8*-GGGAGA^{EXP}. These data suggest that additional, yet-to-be-identified repeat expansions,

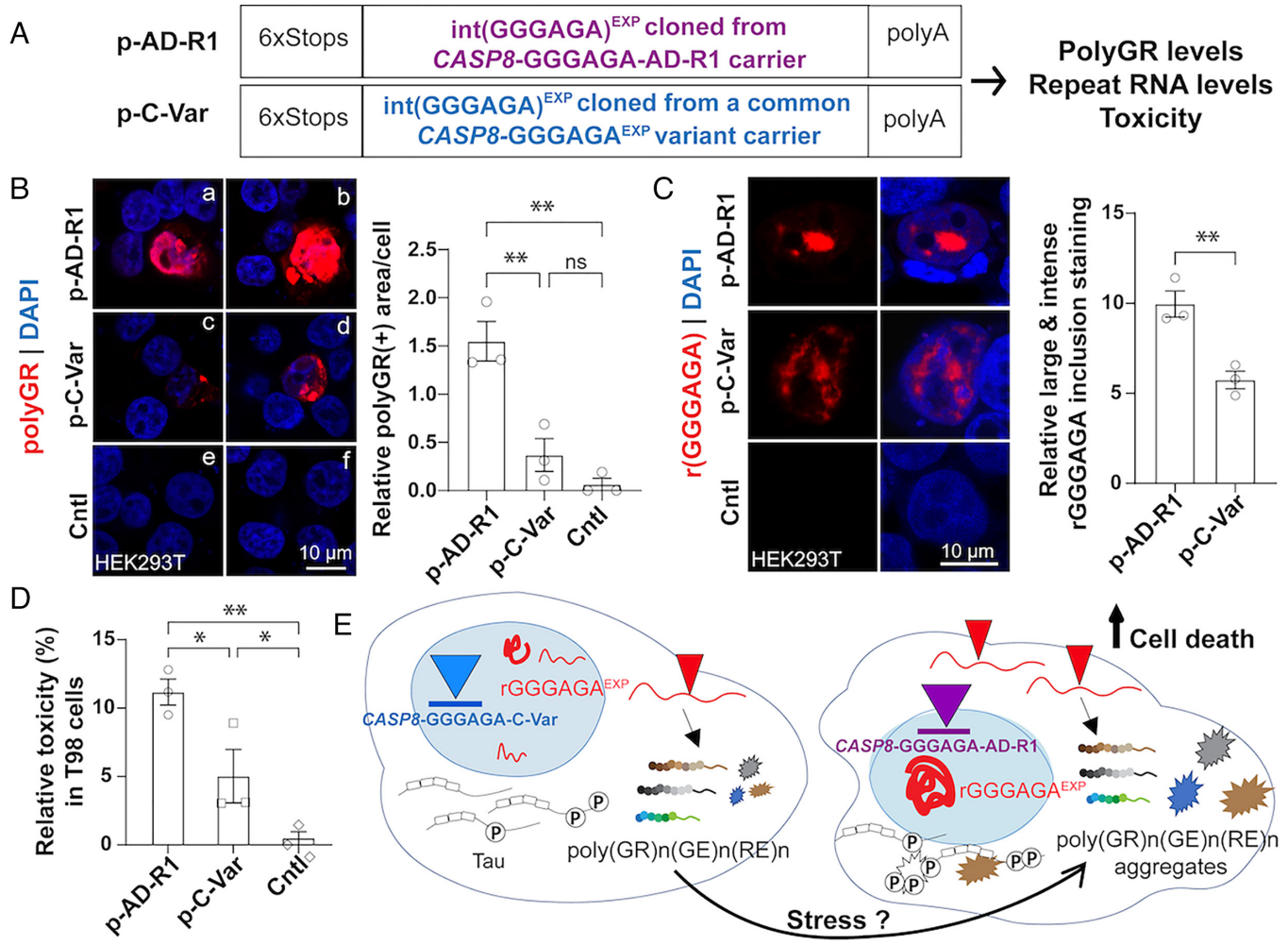

Fig. 6. AD risk variant *CASP8*-GGGAGA-AD-R1 increases RNA and polyGR-containing RAN aggregate levels in cells. (*A*) Schematic diagram of constructs used to express *CASP8*-GGGAGA-AD-R1 (p-AD-R1) and a common *CASP8*-GGGAGA^{EXP} variant (p-C-Var) in transfected cells. (*B*) Example IF images showing a-polyGR staining in HEK293T cells transfected with p-AD-R1 or p-C-Var plasmids and the graph showing quantification of a-polyGR staining (n = 3/group), Cntl = empty vector control. (*C*) Example FISH images showing rGGGAGA^{EXP} RNA staining in HEK293T cells transfected with p-AD-R1, p-C-Var, or empty vector control (Cntl) plasmids and graph showing quantification of GGGAGA RNA inclusions (n = 3/group). (*D*) LDH toxicity assays showing effects of p-AD-R1, p-C-Var, and Cntl minigenes on survival of T98 cells (n = 3/group). (*E*) Model in which stress and allele configuration increase poly(GR)n(RE)n(GE)n and GGGAGA^{EXP} and p-Tau aggregates, which leads to increased AD risk. Data represent mean ± SEM. (*B* and *D*) One-way ANOVA Holm–Sidak's multiple comparisons test, (*C*) unpaired two-tailed *t* test, $^{*}P < 0.05$, $^{**}P < 0.01$.

including potentially other SVA elements, underlie the polyGR pathology found in *CASP8*-GGGAGA^{EXP}(–) AD autopsy brains.

Among the 2,413 AD and control cases, we found *CASP8* GGGAGA expansion of ≥ 47 repeats. Given that the reference genome sequence of the *CASP8* SVA reports an interrupted (GGGAGA) with seven repeats, it is likely that short-read sequencing methods used were not able to accurately sequence the complex repeat expansion at the *CASP8* SVA locus. The identification of pathogenic repeat expansion mutations is challenging and has historically required families with strong histories of disease. dCas9READ combined with RAN protein pathology signatures in patient samples now enable the direct identification of pathogenic repeat expansions from individual patient DNA samples. This approach opens the door for personalized medicine that will allow the direct identification of novel RAN-protein encoding repeat expansion mutations in individual patients and testing the efficacy of RAN-protein lowering drugs including metformin (62). In the future, second generation dCas9READ strategies that use CRISPR/Cas9 variants with more relaxed PAM sequences (63–65) will enable the isolation of additional types of repeat expansion motifs, which will further facilitate efforts to understand the role of the repeatome in biology and disease.

SVA elements have been found at ~3,000 loci in the human genome (66, 67). SVA elements have recently been linked to X-linked dystonia parkinsonism (68), Fukuyama congenital muscular dystrophy (69), and Parkinson's disease (70). Additionally, the *CASP8* SVA-E insertion has been linked with several types of cancers (71, 72). De novo SVA insertions have been shown to contribute to disease and other phenotypes by disrupting gene expression (73–75). Despite growing evidence that SVA elements contribute to disease and phenotypic diversity, there is still limited knowledge about SVA biology. Our work shows that expanded GGGAGA repeats within the *CASP8* SVA-E element produce polymeric proteins that accumulate in AD brains and identify a novel SVA mediated mechanism for neurodegenerative disease.

The *CASP8* GGGAGA^{EXP} shows minimal repeat length variation in brain and blood from two AD cases using bulk genomic DNA and long-range PCR although it is possible that repeat length instability may be detected in some cell populations by small pool PCR. Interruptions are highly variable from case to case. Our data demonstrating that expression of the AD-associated *CASP8*-GGGAGA-AD-R1 variant leads to increased polyGR-RE-GE protein aggregates and RNA inclusions compared to a common *CASP8* GGGAGA^{EXP} variant allele suggests R1 molecular differences in the accumulation of expansion proteins or RNAs may lead to increased AD risk. A potential similarity with other microsatellite expansion diseases is that interruptions within the *CASP8*-

GGGAGA^{EXP} can affect disease penetrance and age of onset (76, 77). In the future, it will be important to better understand the molecular effects of *CASP8*-GGGAGA-AD-R1 and other *CASP8*-GGGAGA^{EXP} variants in disease.

The *CASP8*-SVA-E insertion is a recent human event found in ~56% of the White non-Hispanic population. While tagging SNPs (e.g. rs1035142) are in linkage disequilibrium (r^2 = 0.999) with the *CASP8* SVA insertion, previous GWAS studies would likely have missed the association of the *CASP8*-GGGAGA-AD-R1 variant with AD because the previously used rs1035142 SNP does not distinguish the less common *CASP8*-GGGAGA-AD-R1 variants from the more frequent *CASP8*-SVA-containing alleles. These data highlight that the increased mutability of tandem repeats (78, 79), masks LD with flanking SNPs even within LD blocks, which have no recombination. In SCA2, another microsatellite expansion disorder, CAG repeats with intermediate lengths in *ATXN2* were identified as an ALS risk factor through interactor screens (80). Similar to our *CASP8* study, SCA2 expansions as a modifier of ALS may have been initially missed by GWAS studies because tagging SNPs used in GWAS studies would not have been able to specifically tag *ATXN2* intermediate repeat-length alleles compared to other SCA2 expansion alleles. While our study was focused on three independent White non-Hispanic cohorts the *CASP8*-GGGAGA-AD-R1 variant was also detected in a small cohort of African American AD cases. In future studies, it will be interesting to test whether the *CASP8*-GGGAGA-AD-R1 variant is also associated with AD in other ethnic groups.

The caspase-8 protein plays roles in amyloid metabolism, synaptic plasticity, learning, memory, and the regulation of microglial proinflammatory activation (81, 82). Potential protein loss of function caused by *CASP8*-GGGAGA^{EXP} alleles would be consistent with the LOF effects previously reported for rare *CASP8* protein-coding variants (37). Protein LOF through haploinsufficiency is a common theme in repeat expansion disorders (83). Additionally, the *CASP8* SVA-E allele was previously reported to associate with increased intron 8 retention (39). Both *CASP8* SVA-E tagging SNPs, rs700635 and rs1035142, were previously shown to be strongly associated with increased CASP8 transcript levels in brain tissue and a possible cryptic splicing event upstream of the SVA insertion (39) (GTEx Portal on 12/26/2024). Although not significant, we observed a similar trend toward increased CASP8 transcript levels in a small cohort of *CASP8*-GGGAGA^{EXP}(+) versus *CASP8*-GGGAGA^{EXP}(–) AD cases using end-stage frontal cortex tissue. Additional studies will be required to fully understand the effects of *CASP8*-GGGAGA-AD-R1 and other common *CASP8*-GGGAGA^{EXP} variants on caspase-8 activity and CASP8 developmental and tissue-specific expression patterns. It is possible that splicing alterations including intron retention caused by *CASP8*-GGGAGA-AD-R1 or other common *CASP8*-GGGAGA^{EXP} variants contribute to disease as has been reported in other repeat expansion disorders (84).

The *CASP8*-GGGAGA-AD-R1 expansion variant is strikingly different from other AD risk loci. Most variants that cause monogenic AD (*APP, PSEN1, PSEN2*) or elevate AD risk (*TREM2, ABCA7, SORL1*) with strong effect sizes (OR>3) are relatively infrequent in the general population (<1.5%), except for *APOE4* (OR>3, <20%) (85, 86). In contrast, AD risk alleles with the highest frequencies in the general population [e.g. *ACE (~45%), ECHDC3 (~39%), PTK2B (~37%)*] have low effect sizes (OR<1.2). The *CASP8*-GGGAGA-AD-R1, which has a relatively high OR (2.2) for AD, is also frequently found in 7.5% of AD and 3.6% of controls. These findings, combined with pathological data showing the *CASP8* polymeric proteins detected with locus-specific antibodies accumulate in AD brains carrying both *CASP8*-GGGAGA-AD-R1 and other *CASP8* GGGAGA^{EXP} variants, but not control brains, identify the *CASP8*-GGGAGA repeat as an important and novel risk factor for AD in non-Hispanic Whites.

Materials and Methods

IHC was used to detect a-polyGR(+) aggregates and the unique C-terminal regions expressed from the *CASP8*-GGGAGA^{EXP} mutation in AD autopsy brains. CRISPR deactivated Cas9-based pulldowns were performed using guide RNAs that target GR encoding repeats to enrich and identify candidate repeat expansions express polyGR-containing proteins in AD brains. The *CASP8*-GGGAGA expansions were characterized using short- and long-read sequencing, repeat-primed PCR, and long-range PCR. Case–control association studies were performed to study the frequencies of *CASP8*-GGGAGA^{EXP} variants in AD and control cohorts. Transfected cells were used to study the toxicity of *CASP8*-GGGAGA^{EXP} variants and the production of polyGR-containing RAN proteins from *CASP8*-GGGAGA expansion alleles. Additional details and experimental procedures are provided in *SI Appendix, Materials and Methods*.

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We thank AD patients and their families, the NIH *All of Us* Research Program, Drs. M.S. Swanson and J.D. Cleary for helpful suggestions and funding from the NIH (R37NS040389, RF1NS126536, RF1NS098819, K99AG065511, R00AG065511, and R01NS122766), the Department of Defense (W81XWH2210592), the McKnight Brain Institute, McJunkin Family Foundation, and the Johns Hopkins Alzheimer's Disease Research Center (NIH P30 AG0066507). The UF Neuromedicine Human Brain and Tissue Bank was supported by grants from the National Institute on Aging (NIA) (P30AG066506, P50AG047266). We gratefully acknowledge that DNA samples from the National Centralized Repository for Alzheimer's Disease and Related Dementias (NCRAD), which receives government support under a cooperative agreement grant (U24 AG021886) awarded by the NIA, were used in this study. We thank contributors who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible. Additional details on individual NCRAD contributors can be found in *SI Appendix*.

Author affiliations: [a]Center for NeuroGenetics, College of Medicine, University of Florida, Gainesville, FL 32610; [b]Department of Molecular Genetics & Microbiology, College of Medicine, University of Florida, Gainesville, FL 32610; [c]Genetics Institute, University of Florida, Gainesville, FL 32610; [d]Raymond G. Perelman Center for Cellular and Molecular Therapeutics, The Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA 19104; [e]Department of Pathology & Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104; [f]Department of Pathology, The Johns Hopkins University School of Medicine, Baltimore, MD 21287; [g]Division of Medical Genetics School of Medicine, University of Washington, Seattle, WA 98195; [h]Department of Neurology, MassGeneral Institute for Neurodegenerative Disease, Massachusetts General Hospital, Boston, MA 02114; [i]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455; [j]Department of Pathology, Immunology and Laboratory Medicine, College of Medicine, University of Florida, Gainesville, FL 32610; [k]Center for Translation Research in Neurodegenerative Disease, College of Medicine, University of Florida, Gainesville, FL 32610; [l]Department of Neuroscience, College of Medicine, University of Florida, Gainesville, FL 32610; [m]McKnight Brain Institute, University of Florida, Gainesville, FL 32610; [n]Department of Biostatistics, University of Florida, Gainesville, FL 32611; [o]Ronald M. Loeb Center for Alzheimer's Disease, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029; [p]Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029; [q]Department of Neurology, College of Medicine, University of Florida, Gainesville, FL 32611; and [r]Norman Fixel Institute for Neurological Disease, University of Florida, Gainesville, FL 32608

Author contributions: L.N., T.Z., M.B.C., and L.P.W.R. designed research; L.N., R.A., S.G., L.E.L.R., R.F.T., L.R.B., K.T., and P.N.V. performed research; L.N., P.T.R., J.R.-O., A.M., T.R.C., A. Gaona, O.P., H.B.C., B.L.D., E.T.W., S.P., J.C.T., B.T.H., and L.P.W.R. contributed new reagents/analytic tools; L.N., R.A., S.G., L.E.L.R., R.F.T., L.R.B., P.T.R., T.Z., M.B.C., C.P.K., E.N., H.B.C., A.T.Y., T.E.G., X.Y.L., E.T.W., A.E.R., A. Goate, P.N.V., S.P., J.C.T., B.T.H., and L.P.W.R. analyzed data; and L.N. and L.P.W.R. wrote the paper.

Competing interest statement: T.E.G. is a cofounder and head of scientific advisory board for Andante Biologics. H.B.C. served as a consultant for Janssen Research and Development in 2021 and 2022. A.G. serves on the SAB of Genetech and Muna Therapeutics. B.T.H. serves on the SAB of Dewpoint and has an option for stock. He serves on a scientific advisory board or is a consultant for AbbVie, Alexion, Ambagon, Aprinoia Therapeutics, Arvinas, Avrobio, AstraZenica, Biogen, Bioinsights, BMS, Cure Alz Fund, Cell Signaling, Dewpoint, Latus, Novartis, Pfizer, Sanofi, Sofinnova, Vigil, Violet, Voyager, WaveBreak. T.E.G. owns stock and stock options in Andante Biologics. B.T.H. has an option for stock in Dewpoint. B.T.H. owns stock in Novartis. T.E.G. is an inventor on multiple patents and patent applications that relate to AD therapeutics. L.P.W.R. is an inventor on patents that are related to RAN translation. L.N. is an inventor on patents that are related to RAN translation.

1. K. B. Rajan, J. Weuve, L. L. Barnes, R. S. Wilson, D. A. Evans, Prevalence and incidence of clinically diagnosed Alzheimer's disease dementia from 1994 to 2012 in a population study. *Alzheimers Dement* **15**, 1–7 (2019).
2. R. Cacace, K. Sleegers, C. Van Broeckhoven, Molecular genetics of early-onset Alzheimer's disease revisited. *Alzheimers Dement* **12**, 733–748 (2016).
3. J. Kim, J. M. Basak, D. M. Holtzman, The role of apolipoprotein E in Alzheimer's disease. *Neuron* **63**, 287–303 (2009).
4. S. J. Andrews *et al.*, The complex genetic architecture of Alzheimer's disease: Novel insights and future directions. *EBioMedicine* **90**, 104511 (2023).
5. K. Lobello, J. M. Ryan, E. Liu, G. Rippon, R. Black, Targeting Beta amyloid: A clinical review of immunotherapeutic approaches in Alzheimer's disease. *Int. J. Alzheimers Dis.* **2012**, 628070 (2012).
6. G. P. Morris, I. A. Clark, B. Vissel, Inconsistencies and controversies surrounding the amyloid hypothesis of Alzheimer's disease. *Acta Neuropathol. Commun.* **2**, 135 (2014).
7. H. Paulson, Repeat expansion diseases. *Handb. Clin. Neurol.* **147**, 105–123 (2018).
8. C. Depienne, J. L. Mandel, 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *Am. J. Hum. Genet.* **108**, 764–785 (2021).
9. M. DeJesus-Hernandez *et al.*, Expanded GGGGCC hexanucleotide repeat in noncoding region of C9orf72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256 (2011).
10. A. E. Renton *et al.*, A hexanucleotide repeat expansion in C9orf72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268 (2011).
11. L. Nguyen, J. D. Cleary, L. P. W. Ranum, Repeat-associated non-ATG translation: Molecular mechanisms and contribution to neurological disease. *Annu. Rev. Neurosci.* **42**, 227–247 (2019).
12. T. Zu *et al.*, Non-ATG–initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 260–265 (2011).
13. M. Bañez-Coronel *et al.*, RAN translation in Huntington disease. *Neuron* **88**, 667–677 (2015).
14. T. Zu *et al.*, RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E4968–4977 (2013).
15. L. Nguyen *et al.*, Antibody therapy targeting RAN proteins rescues C9 ALS/FTD phenotypes in C9orf72 mouse model. *Neuron* **105**, 645–662.e611 (2020).
16. T. Zu *et al.*, Metformin inhibits RAN translation through PKR pathway and mitigates disease in C9orf72 ALS/FTD mice. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18591–18599 (2020).
17. R. Gupta *et al.*, The proline/arginine dipeptide from hexanucleotide repeat expanded C9ORF72 inhibits the proteasome. *eNeuro* **4**, ENEURO.0249 (2017).
18. S. Mizielinska *et al.*, C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins. *Science* **345**, 1192–1194 (2014).
19. X. Wen *et al.*, Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death. *Neuron* **84**, 1213–1225 (2014).
20. I. Kwon *et al.*, Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**, 1139–1145 (2014).
21. M. Masnata, S. Salem, A. de Rus Jacquet, M. Anwer, F. Cicchetti, Targeting Tau to treat clinical features of Huntington's disease. *Front. Neurol.* **11**, 580732 (2020).
22. R. Vuono *et al.*, The role of tau in the pathological process and clinical expression of Huntington's disease. *Brain* **138**, 1907–1918 (2015).
23. M. L. Caillet-Boudin *et al.*, Brain pathology in myotonic dystrophy: When tauopathy meets spliceopathy and RNAopathy. *Front. Mol. Neurosci.* **6**, 57 (2014).
24. H. Garcia-Moreno *et al.*, Tau and neurofilament light-chain as fluid biomarkers in spinocerebellar ataxia type 3. *Eur. J. Neurol.* **29**, 2439–2452 (2022).
25. A. P. de Koning, W. Gu, T. A. Castoe, M. A. Batzer, D. D. Pollock, Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **7**, e1002384 (2011).
26. C. Vance *et al.*, Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3. *Brain* **129**, 868–876 (2006).
27. P. N. Valdmanis *et al.*, Three families with amyotrophic lateral sclerosis and frontotemporal dementia with evidence of linkage to chromosome 9p. *Arch. Neurol.* **64**, 240–245 (2007).
28. M. DeJesus-Hernandez *et al.*, Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256 (2011).
29. A. E. Renton *et al.*, A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268 (2011).
30. Y. Ikeda, R. S. Daughters, L. P. Ranum, Bidirectional expression of the SCA8 expansion mutation: One mutation, two genes. *Cerebellum* **7**, 150–158 (2008).
31. N. A. Murphy *et al.*, Age-related penetrance of the C9orf72 repeat expansion. *Sci. Rep.* **7**, 2116 (2017).
32. D. Galimberti *et al.*, Incomplete penetrance of the C9ORF72 hexanucleotide repeat expansions: Frequency in a cohort of geriatric non-demented subjects. *J. Alzheimers Dis.* **39**, 19–22 (2014).
33. P. E. Ash *et al.*, Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* **77**, 639–646 (2013).
34. K. Mori *et al.*, The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339**, 1335–1338 (2013).
35. Y. J. Zhang *et al.*, Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *Nat. Med.* **24**, 1136–1142 (2018).
36. E. Kremmer *et al.*, Rat monoclonal antibodies differentiating between the Epstein-Barr virus nuclear antigens 2A (EBNA2A) and 2B (EBNA2B). *Virology* **208**, 336–342 (1995).
37. J. Rehker *et al.*, Caspase-8, association with Alzheimer's disease and functional analysis of rare variants. *PLoS ONE* **12**, e0185777 (2017).
38. M. DeJesus-Hernandez *et al.*, Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256 (2011).
39. S. N. Stacey *et al.*, Insertion of an SVA-E retrotransposon into the CASP8 gene is associated with protection against prostate cancer. *Hum. Mol. Genet.* **25**, 1008–1018 (2016).
40. All of Us Research Program Genomics, Genomic data in the All of Us Research Program. *Nature* **627**, 340–346 (2024).
41. D. J. Sheskin, *Handbook of Parametric and Non-parametric Procedures* (Chapman and Hall/CRC, 2004).
42. K. M. Green *et al.*, RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nat. Commun.* **8**, 2005 (2017).
43. T. Westergard *et al.*, Repeat-associated non-AUG translation in C9orf72-ALS/FTD is driven by neuronal excitation and stress. *EMBO Mol. Med.* **11**, e9423 (2019).
44. Y. Sonobe *et al.*, Translation of dipeptide repeat proteins from the C9ORF72 expanded repeat is associated with cellular stress. *Neurobiol. Dis.* **116**, 155–165 (2018).
45. W. Cheng *et al.*, C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2alpha phosphorylation. *Nat. Commun.* **9**, 51 (2018).
46. S. K. Tusi *et al.*, The alternative initiation factor eIF2A plays key role in RAN translation of myotonic dystrophy type 2 CCUG*CAGG repeats. *Hum. Mol. Genet.* **30**, 1020–1029 (2021).
47. J. Lytton, M. Westlin, M. R. Hanley, Thapsigargin inhibits the sarcoplasmic or endoplasmic reticulum Ca-ATPase family of calcium pumps. *J. Biol. Chem.* **266**, 17067–17071 (1991).
48. O. Thastrup, P. J. Cullen, B. K. Drobak, M. R. Hanley, A. P. Dawson, Thapsigargin, a tumor promoter, discharges intracellular Ca2+ stores by specific inhibition of the endoplasmic reticulum Ca2(+)-ATPase *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2466–2470 (1990).
49. H. Wang *et al.*, SVA elements: A hominid-specific retroposon family. *J. Mol. Biol.* **354**, 994–1007 (2005).
50. W. W. Cheng *et al.*, C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2α phosphorylation. *Nat. Commun.* **9**, 7076970 (2018).
51. M. Jazurek-Ciesiolka *et al.*, RAN translation of the expanded CAG repeats in the SCA3 disease context. *J. Mol. Biol.* **432**, 166699 (2020).
52. S. Gotoh *et al.*, eIF5 stimulates the CUG initiation of RAN translation of poly-GA dipeptide repeat protein (DPR) in C9orf72 FTLD/ALS. *J. Biol. Chem.* **300**, 105703 (2024).
53. C. N. Cook *et al.*, C9orf72 poly(GR) aggregation induces TDP–43 proteinopathy. *Sci. Transl. Med.* **12**, eabb3774 (2020).
54. B. D. Freibaum *et al.*, GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525**, 129–133 (2015).
55. K. Zhang *et al.*, The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525**, 56–61 (2015).
56. A. Jovicic *et al.*, Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat. Neurosci.* **18**, 1226–1229 (2015).
57. A. B. Loveland *et al.*, Ribosome inhibition by C9ORF72-ALS/FTD-associated poly-PR and poly-GR proteins revealed by cryo-EM. *Nat. Commun.* **13**, 2776 (2022).
58. K. Y. Shi *et al.*, Toxic PRn poly-dipeptides encoded by the C9orf72 repeat expansion block nuclear import and export. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E1111–E1117 (2017), 10.1073/pnas.1620293114.
59. A. Misrani, S. Tabassum, L. Yang, Mitochondrial dysfunction and oxidative stress in Alzheimer's disease. *Front. Aging Neurosci.* **13**, 617588 (2021).
60. B. Uttara, A. V. Singh, P. Zamboni, R. T. Mahajan, Oxidative stress and neurodegenerative diseases: A review of upstream and downstream antioxidant therapeutic options. *Curr. Neuropharmacol.* **7**, 65–74 (2009).
61. W. R. Markesbery, Oxidative stress hypothesis in Alzheimer's disease. *Free Radic. Biol. Med.* **23**, 134–147 (1997).
62. T. Zu *et al.*, Metformin inhibits RAN translation through PKR pathway and mitigates disease in C9orf72 ALS/FTD mice. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18591–18599 (2020).
63. R. T. Walton, K. A. Christie, M. N. Whittaker, B. P. Kleinstiver, Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290-+ (2020).
64. B. P. Kleinstiver *et al.*, Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33**, 1293–1298 (2015).
65. B. P. Kleinstiver *et al.*, Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
66. C. Stewart *et al.*, A comprehensive map of mobile element insertion polymorphisms in humans. *PLoS Genet.* **7**, e1002236 (2011).
67. A. L. Pfaff, L. M. Singleton, S. Koks, Mechanisms of disease-associated SINE-VNTR-Alus. *Exp. Biol. Med. (Maywood)* **247**, 756–764 (2022).
68. D. C. Bragg *et al.*, Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in TAF1. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E11020–E11028 (2017).
69. M. Watanabe *et al.*, Founder SVA retrotransposal insertion in Fukuyama-type congenital muscular dystrophy and its origin in Japanese and Northeast Asian populations. *Am. J. Med. Genet. A* **138**, 344–348 (2005).
70. A. L. Pfaff, V. J. Bubb, J. P. Quinn, S. Koks, Reference SVA insertion polymorphisms are associated with Parkinson's disease progression and differential gene expression. *NPJ Parkinsons Dis.* **7**, 44 (2021).
71. S. N. Stacey *et al.*, New basal cell carcinoma susceptibility loci. *Nat. Commun.* **6**, 6825 (2015).
72. S. N. Stacey *et al.*, Insertion of an SVA-E retrotransposon into the CASP8 gene is associated with protection against prostate cancer. *Hum. Mol. Genet.* **25**, 1008–1018 (2016).
73. N. Kamitaki *et al.*, A sequence of SVA retrotransposon insertions in ASIP shaped human pigmentation. *Nat. Genet.* **56**, 1583–1591 (2024), 10.1038/s41588-024-01841-4.
74. M. Taniguchi-Ikeda *et al.*, Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy. *Nature* **478**, 127–131 (2011).
75. T. Aneichyk *et al.*, Dissecting the causal mechanism of X-linked dystonia-parkinsonism by integrating genome and transcriptome assembly. *Cell* **172**, 897–909.e821 (2018).
76. B. A. Perez *et al.*, CCG*CGG interruptions in high-penetrance SCA8 families increase RAN translation and protein toxicity. *EMBO Mol. Med.* **13**, e14095 (2021).
77. G. E. B. Wright *et al.*, Interrupting sequence variants and age of onset in Huntington's disease: Clinical implications and emerging therapies. *Lancet Neurol.* **19**, 930–939 (2020).
78. I. Mitra *et al.*, Patterns of de novo tandem repeat mutations and their role in autism. *Nature* **589**, 246–250 (2021).
79. A. Kong *et al.*, Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471–479 (2012).
80. A. C. Elden *et al.*, Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* **466**, 1069–1075 (2010).
81. B. Tummers, D. R. Green, Caspase-8: Regulating life and death. *Immunol. Rev.* **277**, 76–89 (2017).
82. L. Pellegrini, B. J. Passer, M. Tabaton, J. K. Ganjei, L. D'Adamio, Alternative, non-secretase processing of Alzheimer's beta-amyloid precursor protein during apoptosis by caspase-6 and -8. *J. Biol. Chem.* **274**, 21011–21016 (1999).
83. I. Malik, C. P. Kelley, E. T. Wang, P. K. Todd, Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat. Rev. Mol. Cell Biol.* **22**, 589–607 (2021).
84. ŁJ. Sznajder *et al.*, Intron retention induced by microsatellite expansions as a disease biomarker. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4234–4239 (2018).
85. I. de Rojas *et al.*, Common variants in Alzheimer's disease and risk stratification by polygenic risk scores. *Nat. Commun.* **12**, 3417 (2021).
86. I. E. Jansen *et al.*, Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).