"Protein aggregates" contain RNA and DNA, entrapped by misfolded proteins but largely rescued by slowing translational elongation

Robert J. Shmookler Reis1,2,3 | Ramani Atluri2 | Meenakshisundaram Balasubramaniam2 | Jay Johnson3 | Akshatha Ganne3 | Srinivas Ayyadevara1,2

1Central Arkansas Veterans Healthcare System, Little Rock, AR, USA
2Department of Geriatrics, University of Arkansas for Medical Sciences, Little Rock, AR, USA
3Bioinformatics Program, University of Arkansas for Medical Sciences and University of Arkansas at Little Rock, Little Rock, AR, USA

Correspondence
Robert J. Shmookler Reis and Srinivas Ayyadevara, Central Arkansas Veterans Healthcare System, Little Rock, AR 72205, USA.
Emails: rjsr@uams.edu (R. J. S. R.); ayyadevarasrinivas@uams.edu (S. A.)

Funding Information
U.S. Department of Veterans Affairs, Grant/Award Number: 2 I01 BX001655 and IK6 BX004851; National Institute on Aging, Grant/Award Number: 1 R01 AG062254-01 and 2 P01 AG012411

Abstract
All neurodegenerative diseases feature aggregates, which usually contain disease-specific diagnostic proteins; non-protein constituents, however, have rarely been explored. Aggregates from SY5Y-APPsw neuroblastoma, a cell model of familial Alzheimer’s disease, were crosslinked and sequences of linked peptides identified. We constructed a normalized “contactome” comprising 11 subnetworks, centered on 24 high-connectivity hubs. Remarkably, all 24 are nucleic acid-binding proteins. This led us to isolate and sequence RNA and DNA from Alzheimer’s and control aggregates. RNA fragments were mapped to the human genome by RNA-seq and DNA by ChIP-seq. Nearly all aggregate RNA sequences mapped to specific genes, whereas DNA fragments were predominantly intergenic. These nucleic acid mappings are all significantly nonrandom, making an artifactual origin extremely unlikely. RNA (mostly cytoplasmic) exceeded DNA (chiefly nuclear) by twofold to fivefold. RNA fragments recovered from AD tissue were ~1.5-to 2.5-fold more abundant than those recovered from control tissue, similar to the increase in protein. Aggregate abundances of specific RNA sequences were strikingly differential between cultured SY5Y-APPsw glioblastoma cells expressing APOE3 vs. APOE4, consistent with APOE4 competition for E-box/CLEAR motifs. We identified many G-quadruplex and viral sequences within RNA and DNA of aggregates, suggesting that sequestration of viral genomes may have driven the evolution of disordered nucleic acid-binding proteins. After RNA-interference knockdown of the translational-procession factor EEF2 to suppress translation in SY5Y-APPsw cells, the RNA content of aggregates declined by >90%, while reducing protein content by only 30% and altering DNA content by ≤10%. This

Abbreviations: AD, Alzheimer’s disease; AMC, age-matched controls; APOE, APOE, apolipoprotein E (protein, gene); Aβ1-42, Amyloid beta (residues 1–42); ChIP-seq, Chromatin ImmunoPrecipitation - sequencing; CLEAR, Coordinated Lysosomal Expression and Regulation, a binding motif; DNA, deoxyribonucleic acid; EEF2, EEF2, eukaryotic elongation factor 2 (protein, gene); G4BP, G-quadruplex binding protein; MAP1A, microtubule-associated protein 1A; NA, nucleic acids; NS, not significant (referring to a difference between groups); OE, overexpressing; PTM, post-translational modification; RNA, ribonucleic acid; TF, transcription factor.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Aging Cell published by Anatomical Society and John Wiley & Sons Ltd. Published 2021. This article is a U.S. Government work and is in the public domain in the USA

wileyonlinelibrary.com/journal/acel | 1 of 20
1 | INTRODUCTION

Protein aggregation increases inexorably with aging in all animal species and in all tissues examined (Ayyadevara, Balasubramaniam, Johnson, et al., 2016; Ayyadevara, Balasubramaniam, Suri, et al., 2016; Brignull et al., 2007; Cohen et al., 2009; David et al., 2010; Dillin & Cohen, 2011; Reis-Rodrigues et al., 2012). Specific aggregate components, diagnostic for each human neurodegenerative disease, are thought to play causal roles because their pathology-associated mutation and/or overexpression are sufficient to confer heritable neuropathy in human pedigrees and in transgenic-animal models (Bandyopadhyay et al., 2007; Dillin & Cohen, 2011; Li et al., 2013; Miller et al., 2010; Rooiveldt et al., 2009). Proteins that require structural flexibility often incorporate disordered regions, thus rendering them vulnerable to aggregation; other proteins are only susceptible to aggregation after oxidation or specific post-translational modifications (Ayyadevara et al., 2015, 2017; Ayyadevara, Balasubramaniam, Parcon, et al., 2016).

We recently developed improved click-chemistry crosslinking reagents and analytical software to identify adjacent proteins in aggregates, based on peptide-peptide crosslinking, and we applied it to define the protein-adherence network, or “contactome”, of aggregates. We began with total, sarkosyl-insoluble aggregates isolated from SY5Y-APP-plied it to define the protein-adherence network, or “contactome”, in aggregates, based on peptide-peptide crosslinking, and we applied reagents and analytical software to identify adjacent proteins (Parcon, et al., 2016). Specific aggregate components, diagnostic for each human neurodegenerative disease, are thought to play causal roles because their pathology-associated mutation and/or overexpression are sufficient to confer heritable neuropathy in human pedigrees and in transgenic-animal models (Bandyopadhyay et al., 2007; Dillin & Cohen, 2011; Li et al., 2013; Miller et al., 2010; Rooiveldt et al., 2009). Proteins that require structural flexibility often incorporate disordered regions, thus rendering them vulnerable to aggregation; other proteins are only susceptible to aggregation after oxidation or specific post-translational modifications (Ayyadevara et al., 2015, 2017; Ayyadevara, Balasubramaniam, Parcon, et al., 2016).

We recently developed improved click-chemistry crosslinking reagents and analytical software to identify adjacent proteins in aggregates, based on peptide-peptide crosslinking, and we applied it to define the protein-adherence network, or “contactome”, of aggregates. We began with total, sarkosyl-insoluble aggregates isolated from SY5Y-APP-plied it to define the protein-adherence network, or “contactome”, in aggregates, based on peptide-peptide crosslinking, and we applied reagents and analytical software to identify adjacent proteins (Parcon, et al., 2016). Specific aggregate components, diagnostic for each human neurodegenerative disease, are thought to play causal roles because their pathology-associated mutation and/or overexpression are sufficient to confer heritable neuropathy in human pedigrees and in transgenic-animal models (Bandyopadhyay et al., 2007; Dillin & Cohen, 2011; Li et al., 2013; Miller et al., 2010; Rooiveldt et al., 2009). Proteins that require structural flexibility often incorporate disordered regions, thus rendering them vulnerable to aggregation; other proteins are only susceptible to aggregation after oxidation or specific post-translational modifications (Ayyadevara et al., 2015, 2017; Ayyadevara, Balasubramaniam, Parcon, et al., 2016).

We recently developed improved click-chemistry crosslinking reagents and analytical software to identify adjacent proteins in aggregates, based on peptide-peptide crosslinking, and we applied it to define the protein-adherence network, or “contactome”, of aggregates. We began with total, sarkosyl-insoluble aggregates isolated from SY5Y-APP-plied it to define the protein-adherence network, or “contactome”, in aggregates, based on peptide-peptide crosslinking, and we applied reagents and analytical software to identify adjacent proteins (Parcon, et al., 2016). Specific aggregate components, diagnostic for each human neurodegenerative disease, are thought to play causal roles because their pathology-associated mutation and/or overexpression are sufficient to confer heritable neuropathy in human pedigrees and in transgenic-animal models (Bandyopadhyay et al., 2007; Dillin & Cohen, 2011; Li et al., 2013; Miller et al., 2010; Rooiveldt et al., 2009). Proteins that require structural flexibility often incorporate disordered regions, thus rendering them vulnerable to aggregation; other proteins are only susceptible to aggregation after oxidation or specific post-translational modifications (Ayyadevara et al., 2015, 2017; Ayyadevara, Balasubramaniam, Parcon, et al., 2016).

We recently developed improved click-chemistry crosslinking reagents and analytical software to identify adjacent proteins in aggregates, based on peptide-peptide crosslinking, and we applied it to define the protein-adherence network, or “contactome”, of aggregates. We began with total, sarkosyl-insoluble aggregates isolated from SY5Y-APP-plied it to define the protein-adherence network, or “contactome”, in aggregates, based on peptide-peptide crosslinking, and we applied reagents and analytical software to identify adjacent proteins (Parcon, et al., 2016). Specific aggregate components, diagnostic for each human neurodegenerative disease, are thought to play causal roles because their pathology-associated mutation and/or overexpression are sufficient to confer heritable neuropathy in human pedigrees and in transgenic-animal models (Bandyopadhyay et al., 2007; Dillin & Cohen, 2011; Li et al., 2013; Miller et al., 2010; Rooiveldt et al., 2009). Proteins that require structural flexibility often incorporate disordered regions, thus rendering them vulnerable to aggregation; other proteins are only susceptible to aggregation after oxidation or specific post-translational modifications (Ayyadevara et al., 2015, 2017; Ayyadevara, Balasubramaniam, Parcon, et al., 2016).

We recently developed improved click-chemistry crosslinking reagents and analytical software to identify adjacent proteins in aggregates, based on peptide-peptide crosslinking, and we applied it to define the protein-adherence network, or “contactome”, of aggregates. We began with total, sarkosyl-insoluble aggregates isolated from SY5Y-APP-plied it to define the protein-adherence network, or “contactome”, in aggregates, based on peptide-peptide crosslinking, and we applied reagents and analytical software to identify adjacent proteins (Parcon, et al., 2016). Specific aggregate components, diagnostic for each human neurodegenerative disease, are thought to play causal roles because their pathology-associated mutation and/or overexpression are sufficient to confer heritable neuropathy in human pedigrees and in transgenic-animal models (Bandyopadhyay et al., 2007; Dillin & Cohen, 2011; Li et al., 2013; Miller et al., 2010; Rooiveldt et al., 2009). Proteins that require structural flexibility often incorporate disordered regions, thus rendering them vulnerable to aggregation; other proteins are only susceptible to aggregation after oxidation or specific post-translational modifications (Ayyadevara et al., 2015, 2017; Ayyadevara, Balasubramaniam, Parcon, et al., 2016).

2.2 | Quantitation of aggregate nucleic acids from AD vs. control hippocampus, or human glioma cells

We isolated total sarkosyl-insoluble aggregates from hippocampal tissue of individuals diagnosed with Alzheimer’s disease (AD) and confirmed by histopathological markers, and from age-matched controls (AMC) without dementia or AD-diagnostic markers (amyloid deposits or hyperphosphorylated tau). From equal initial weights of hippocampus, quantified recoveries of nucleic acids increased in AD aggregates, over those in controls, by 1.5- to 2-fold for DNA, and ~twofold for RNA (Figure 2A,B). These elevations did not differ significantly from the difference in protein content of total aggregates, which was ~60% higher in AD than in controls, in close agreement with previous results (Ayyadevara, Balasubramaniam, Parcon, et al., 2016). Among normal controls, there was fourfold to sixfold more RNA than DNA in total sarkosyl-insoluble aggregates (p < 1E-5), regardless of the methods used for separation and quantitation (see Experimental Procedures). For AD samples, nucleic acid recoveries were higher and more variable, with roughly twice as much RNA as DNA (Figure 2B).

Apolipoprotein E (APOE) gene alleles are the leading genetic risk factors for AD, with at least fourfold increased AD risk for each APOE ε4 allele (abbreviated APOE4, ε4, or E4), and increased severity of aggregate-associated neuropathology for AD carriers of APOE4 alleles (Neu et al., 2017; Parker et al., 2005). We recently reported that the concerted transcription of autophagy genes is disrupted in the human glioblastoma cell line T98G, when overexpressing an APOE4 transgene rather than APOE3 (Parcon et al., 2018). To assess whether the greater nucleic acid content of aggregates in AD vs. AMC hippocampi may reflect the disruption of autophagy in AD, we separately analyzed aggregates from T98G cells that overexpress...
FIGURE 1 The “aggregate contactome” of proteins isolated from SY5Y-APPsw human neuroblastoma cells, an in vitro model of familial AD. The contactome was generated from proteomic data for cross-linked peptide pairs in sarkosyl-insoluble aggregates, using a modified version of X-link Identifier (Balasubramaniam et al., 2019; Du et al., 2011), requiring ≥10 spectral hits per protein observed in at least 2 of 3 replicate crosslinking experiments. Hits were normalized to hub length (amino acids in the most abundant isoform). Red circles highlight central hubs with 5 or more large-hub interactors; green circles show smaller hub-connectors, which join major hubs not otherwise connected. Other proteins of interest are indicated by dashed gray circles.
APOE ε3 or ε4 from transgenes. The DNA content of T98G/E4 aggregates was about twice that of T98G/E3 aggregates (Figure 2C; p < 0.001), whereas their RNA content declined a little (<15%, N. S.) with the APOE4 allele overexpressed.

Most neuropathic aggregates are cytoplasmic, but may also be nuclear or extracellular. When we separated nuclei from cytoplasm of T98G cells prior to aggregate isolation, similar amounts of aggregate protein were recovered from each fraction. However, nuclear aggregates contained mostly DNA and only 40% as much RNA, while cytoplasmic aggregates contained ~10-fold more RNA than DNA (Figure 2D).

### 2.3 Sequencing data for aggregate nucleic acids

To assess whether DNA and RNA fragments in aggregates are a random sampling from the genome and transcriptome, respectively, these nucleic acids were separately extracted from pooled aggregate preparations from either AD or age-matched control (AMC) individuals (APOE ε3/ε4 heterozygotes; 3 subjects per group), and their sequences determined (UT Southwestern Genomics Core, Dallas TX). DNA fragments were analyzed using a ChIP-seq protocol, suited to detection of site specificity, and were then mapped to the human genome. RNA fragments underwent a dual screening, comprising a test of peak significance (similar to ChIP-seq) followed by RNA-seq analysis of differential abundance, and mapping to the human genome.

Of the 38 loci that showed significant DNA read peaks in at least one pool (each p < 10⁻⁵ that the reads actually came from a uniform distribution), 25 reached that threshold in AD vs. 28 in AMC tissue, and 26 in T98G cells carrying APOE3 vs. 30 in APOE4 cells (Table 1, Supplementary Table S1). Each group contains 17–20 peak loci with p < 10⁻⁶, and one locus with p < 10⁻⁵⁰. Considering that a random representation of DNA reads would display a flat (uniform) distribution, these data provide decisive evidence that aggregate DNA fragments are not random, but most likely reflect specific binding sites for proteins that ensnared them into aggregates. Only 5 of these 38 DNA peaks (13%) mapped to known genes or ORFs: RPS5-857K21.4, LINC00486, DUX4L26, MAMDC2-AS1, and ROCK1P1. The other 33 peaks mapped to intergenic regions. Of the 24 chromosomes represented, Y had the most DNA reads (47–98 reads in each of 6 peaks), followed by chromosomes 4 (3 peaks of 48–53 reads), 1 (2 peaks of 57–65 reads), 10 (2 peaks of 35–45 reads), 16 (2 peaks of 26–35 reads), 17 (2 peaks of 24–37 reads), and 21 (4 peaks of 9–17 reads). Neither the numbers of DNA peaks or reads differed significantly between the groups compared (AD vs. AMC, or E3 vs. E4), with the sole exception of a Y-chromosome peak mapping at 26,638,004−26,638,595, which yielded only 5 reads in aggregates from AD, but 59 for AMC (Chi² p < 0.005) and a range of 52–63 reads for the 3 non-AD samples.
| Region | Chrom. Region | Length | 5' Gene | 5' distance | 3' Gene | 3' distance | AD Peak p-value | AMC Peak p-value | Transcribed processed pseudogene, also in all aggregate RNAs |
|--------|---------------|--------|---------|-------------|---------|-------------|----------------|-----------------|----------------------------------------------------------|
| 1      | 633735..643000 | 566    | RP5-857K12.3 | 34236 | RP5-857K21.4 | 0          | 1             | 4               | 4.02             | 4.06                                                        |
| 2      | 132920384..132937903 | 569    | MIR8078 | 421 | ROCK1P1 | 17       | 2             | 3.07            | 3.07             | MAM Domain Containing 2 (antisense strand)                  |
| 3      | 756395..756953 | 566    | RP5-857K21.3 | 2163836 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Double Homeobox 4-Like                                      |
| 4      | 49771147..49771912 | 566    | RP11-341K13.1 | 0 | UNCO0046 | 0          | 1             | 4               | 4.11             | Long Intergenic Non-Protein Coding RNA (antisense strand)  |
| 5      | 89830806..89831378 | 573    | 55'-RNA | 230125 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Long Intergenic Non-Protein Coding RNA (antisense strand)  |
| 6      | 143202361..143202903 | 543    | RP5-857K21.3 | 2163836 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Double Homeobox 4-Like                                      |
| 7      | 756683..7566933 | 566    | RP5-857K21.3 | 2163836 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Double Homeobox 4-Like                                      |
| 8      | 49171147..49771912 | 566    | RP11-341K13.1 | 0 | UNCO0046 | 0          | 1             | 4               | 4.11             | Long Intergenic Non-Protein Coding RNA (antisense strand)  |
| 9      | 70038006..70038625 | 562    | RN7SL570P | 1572274 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Long Intergenic Non-Protein Coding RNA (antisense strand)  |
| 10     | 42070370..42070986 | 562    | RN7SL570P | 1572274 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Long Intergenic Non-Protein Coding RNA (antisense strand)  |
| 11     | 12345678..12345678 | 562    | RN7SL570P | 1572274 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Long Intergenic Non-Protein Coding RNA (antisense strand)  |
| 12     | 35645678..35645678 | 562    | RN7SL570P | 1572274 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Long Intergenic Non-Protein Coding RNA (antisense strand)  |
| 13     | 35645678..35645678 | 562    | RN7SL570P | 1572274 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Long Intergenic Non-Protein Coding RNA (antisense strand)  |
| 14     | 35645678..35645678 | 562    | RN7SL570P | 1572274 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Long Intergenic Non-Protein Coding RNA (antisense strand)  |
| 15     | 35645678..35645678 | 562    | RN7SL570P | 1572274 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Long Intergenic Non-Protein Coding RNA (antisense strand)  |
| 16     | 35645678..35645678 | 562    | RN7SL570P | 1572274 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Long Intergenic Non-Protein Coding RNA (antisense strand)  |
| 17     | 35645678..35645678 | 562    | RN7SL570P | 1572274 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Long Intergenic Non-Protein Coding RNA (antisense strand)  |
| 18     | 35645678..35645678 | 562    | RN7SL570P | 1572274 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Long Intergenic Non-Protein Coding RNA (antisense strand)  |
| 19     | 35645678..35645678 | 562    | RN7SL570P | 1572274 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Long Intergenic Non-Protein Coding RNA (antisense strand)  |
| 20     | 35645678..35645678 | 562    | RN7SL570P | 1572274 | CH17-333M13.2 | 104276 | 2             | 4               | 4.11             | Long Intergenic Non-Protein Coding RNA (antisense strand)  |
TABLE 1 (Continued)

Note: All listed peaks differed significantly from a uniform distribution at p < 0.05 relative to AMC. Peak counts were not corrected for the 1.6-fold higher protein and DNA recovery from AD relative to AMC hippocampus, since all reflect normalized data from 1 µg DNA.

RNA differed from DNA in several important respects. Most notably, 81% of RNA peaks mapped to known or putative genes, vs. only 13% of DNA peaks. (Table 2, Supplementary Tables S2, S3; note that intergenic RNA peaks were omitted to conserve space.) The 49 “within-peak” genes include 39 (80%) that differed significantly in read count between AD and AMC at p < 0.001 (2-tailed Fisher exact tests), vs. 1 of 38 (2.6%) for DNA. It is noteworthy that 30 significantly differential genes were more abundant in AD, while only 9 (23%) were relatively enriched in AMC. This 3:1 bias is on top of the ~1.8-fold higher abundance of RNA in AD aggregates (Figure 2A), since all counts were normalized to the source library. Among T98G glioblastoma genes (Supplementary Table S3), 54 of 59 RNA peaks (92%) differ between APOE3 and APOE4 at p < 0.0001, in marked contrast to DNA peaks of which none were significantly differential.

Aggregate RNA reads that were substantially more abundant in AD than controls (Chi-square or 2-tailed Fisher exact p < 0.001) include two uncharacterized transcripts on chromosome 21, enriched 17-fold and 9-fold beyond other RNAs in AD; ribosomal protein/RPS29, 4.6-fold; RNase-P/RPPH1_2, 3.7-fold; nuclear RNA/SNORD3A and long noncoding RNA/LINC00486, enriched 3.5- and 3.4-fold; signal-recognition-particle RNAs (SRP_138 and RN7SK), 3.2- and 2.7-fold; mitochondrial RNase P1/RMRP1, 2.6-fold; karyopherin/KPNA4, 2.3-fold; amyloid precursor protein/APP, 2.1-fold, and SERCA2/ATP2A2, 1.9-fold (Table 2). It is noteworthy that 2 of the 5 genes identified in aggregate DNA, LINC00486 and RP5-857K21.4, were also among the AD-enriched transcripts in aggregates, and 9 of the 39 genes (23%) enriched in AD aggregate RNA, relative to AMC, encode proteins that were also enriched in AD-specific aggregates (Ayyadevara, Balasubramaniam, Parcon, et al., 2016) (bold font in the rightmost column of Table 2).

Because we had observed roughly twice as much DNA in aggregates isolated from glioblastoma cells overexpressing APOE4, as in identical cells expressing APOE3, we asked whether any particular loci or genes were differentially represented in their aggregates. DNA read counts from E3 and E4 aggregates were in fact quite similar for all DNA loci sequenced (Supplementary Table S1), whereas RNA sequencing data (listed in Supplementary Table S3) show striking increases in aggregate-entrapped RNA transcripts isolated from APOE3-overexpressing (OE) cells, relative to isogenic cells overexpressing APOE4. For 53 of the 59 genes that were confidently identified within fragment alignment peaks, the read count in APOE4-OE cells differed significantly from APOE3-OE cells at Fisher exact p < 10^-4, with E3/E4 ratios ranging from 1.7-13.9. Only one gene appeared to be more abundant in the presence of excess APOE4, a long noncoding RNA for which there were too few reads to attain significance. This bias is consistent with evidence that APOE4 protein competes with transcription factor TFEB for the ~400 DNA binding sites containing the CLEAR motif, most of which drive expression of proteins involved in autophagy/lysosome functions (Sardiello, 2016).

Differential RNA-fragment abundances in glioblastoma aggregates, in which only the APOE allele differs between cell lines, tend to be highly significant (53 of 59 have p < 0.0001) and comprise an interesting set. Examples include α-enolase (E3/E4 = 1.85), MHC-II YBX1...
(3.8), scaRNA2 (2.3), histones H2B (2.8) and H1 (6.7), HSP60 (1.6), HSP90-A1 (1.8) and -B1 (1.9), HSP-A8 (2.0) and HSP-B1 (3.4), IGF-BP3 (5.4) and -B5 (2.0), NCL (2.0), prothymosin α (3.2), SPARC (2.4), nucleophosmin (2.5), RACK1 (3.0), 75K small nuclear RNA (6.2), EEF1-A1 (2.2), β-actin (3.2), peptidylprolyl isomerase A (2.6), collagen 1A2 (4.1), vimentin (4.2), CD44 (2.7), coflin 1 (3.6), GAPDH (2.9), α tubulin (2.9), RNAse P component H1 (9.2), ribosomal proteins RPS2 (4.2) and RPS29 (13.9), 7SL RNA 2 (8.7), β2-microglobulin (4.1), annexin A2 (3.4), pyruvate kinase M (2.4), profilin 1 (3.7), γ-actin (3.4), APOE (3.6), ferritin light chain 1 (3.0), galectin 1 (5.0), and filamin A (2.1). With lesser significance, we find synapsin 3 (E3/E4 = 3.5; \( p < 0.0002 \)), sequestosome 1/SQSTM1 (1.3; \( p < 0.004 \)) and vimentin antisense (15; \( p < 0.002 \)).

In both the direction and magnitude of the RNA-abundance shift, the influence of Alzheimer’s disease was less consistent and so appeared less pronounced on average, than that of the APOE allele. This is almost certainly due to genetic and environmental variance among AD and AMC subjects (Ayyadevara, Balasubramaniam, Parcon, et al., 2016), in contrast to the single transgene that distinguishes T98G/E3 from T98G/E4 cells. Because all human subjects considered in the present comparison were APOE3/E4 heterozygotes, the AD effect could not have arisen from a difference in APOE genotypes. The prevailing reduction in RNA content of E4 aggregates, for the most differentially expressed genes, may reflect transcriptional suppression of TFEB targets by APOE4 (Parcon et al., 2018), rather than an impact of the APOE allele on aggregation per se.

Mapping the RNA transcripts to the human genome revealed a remarkable cluster of at least 20 intergenic loci in a relatively silent segment (21p11.2–21p12) of the chromosome 21 short arm (Supplementary Figure S1). While these loci are not differentially represented for the most part, either between AD and AMC or between APOE3 and APOE4, they include 2 loci with the highest AD/AMC ratios we observed, 9.1 and 17.2 (each \( \chi^2 p < 10^{-6} \), Table 2).

2.4 | Annotation enrichment meta-analysis of RNA fragments in aggregates

Although we had expected the RNA fragments embedded in aggregates to comprise a random selection from the transcriptome, gene ontology and pathway term enrichment analysis (functional-annotation clustering in DAVID\textsuperscript{TM}, http://david.ncifcrf.gov) revealed highly significant enrichment for specific groups of RNAs. Focusing on genes with RNA reads that map to significant peaks and are differentially abundant in T98G/E3 vs. T98G/E4 glioblastoma cells, DAVID meta-analysis revealed highly significant enrichment clusters for gene annotations relating to [extracellular exosome + actylation + phosphoprotein + nucleus, actylation + poly(A) binding], [Ubl conjugation + cadherin binding + cell-cell adherens junction], [glycoprotein binding + protein stabilization], and [myelin sheath + unfolded protein response + protein refolding + stress response + chaperone] (Table 3A).

Among genes with well-mapped reads that differ significantly between aggregates from AD vs. AMC, the most enriched clusters include [intracellular ribonucleoprotein complex + methylation + Ubl conjugation + poly(A) RNA binding + actylation], [extracellular matrix + chaperone + ATPase activity], [nucleosplasm + nucleus], and [myelin sheath + unfolded protein binding + chaperone] (Table 3B). The very existence of these clusters, and their marked overlap between meta-analyses derived from gene lists of very different origin (aggregates from cultured glioblastoma cells vs. human hippocampi) despite only 11 common members, suggests that the underlying aggregate-RNA fragments are strikingly nonrandom in nature. The specific annotation terms that were most enriched (Table 3C) are likely to reflect the nature of proteins that coalesce in AD and AD-model aggregates, which include terms (fold enrichment) such as protein refolding (70), MHC class II protein complex binding (62), oxidation (56), amyloidosis (43), response to unfolded protein (31), glycoprotein binding (27), intracellular ribonucleoprotein complex (18), unfolded protein binding (16), and neurodegeneration (9).

2.5 | What mechanisms account for RNA and DNA fragments co-aggregating with proteins?

What does the inclusion of DNA and RNA fragments imply about aggregates or the mechanism of aggregation? Clearly, there are proteins in aggregates that evolved to bind both nucleic acids and proteins. RNA assumes many transient structures constrained chiefly by its duplex regions, which form A-helices. DNA, in addition to its repertoire of relatively stable structures (A-, B- and Z-duplex helices, triplex, and G-quadruplex forms), in the course of replication and transcription can adopt as wide a range of single-stranded structures as RNA. Affinity for nucleic acids, as well as the protein constituents of multimeric RNA- and DNA-binding complexes, may require protein structures that are at least partially disordered (Zhang et al., 2013) and/or are highly polar, which in turn may favor aggregation (Babu, 2016; Kovacech et al., 2010). DNA-binding proteins include histones; high-mobility-group (HMG) proteins; constituents of DNA replication, transcription, and repair complexes (e.g. topoisomerases, helicases and polymerases; transcription factors, co-factors, and repressors); and proteins that stabilize or remodel chromatin (Figure 1) (Li et al., 2006; Mitchell & Tjian, 1989; Stoyanova et al., 2009; Wade, 2001). A key feature shared by many DNA-binding proteins, in addition to structural instability, is an excess of positively charged residues—allowing formation of electrostatic bonds to the negatively charged phosphates that link DNA-backbone sugars.

RNA-binding proteins include splicing factors, translational initiation and elongation factors, ribosomal and associated proteins (e.g., refolding chaperones), signal recognition particles, and proteins involved in the processing and functions of noncoding RNAs (Figure 1) (Castello et al., 2012; Glisovic et al., 2008; Hentze et al., 2018; Turner & Hodson, 2012). There are also diverse proteins that bind both RNA and DNA—including RNA polymerases and other transcription-complex components, RNA/DNA helicases, and TAR DNA-binding protein (TDP43/TADBP) (Gao et al., 2019; Hudson & Ortlund, 2014; Kobren & Singh, 2019; Nikpour & Salavati, 2019;
| Chrom. Region | Length | 5' dist. | 3' dist. | 5' gene | 3' gene | AD reads | AMC reads | Chi^2_p | Transcribed/encoded protein |
|---------------|--------|----------|----------|---------|---------|----------|-----------|--------|----------------------------|
| 1             | 433888..434164 | 277     | 277      | RP5-857K21.3 | RP5-857K21.4 | 0        | 515       | 0.0001 | Transcribed/processed pseudogene |
| 1             | 2281859..2326263 | 405     | 405      | SKI     | MORN1   | 36899   | 396       | 0.0001 | Splicing factor 6 |
| 1             | 109100211..109100485 | 275     | 275      | SCARNA2 | RP11-11.N7.5 | 8277    | HNRNPU    | 0.38   | Small Cajal-body RNA (scRNA) guides mod'rn of siRNAs |
| 1             | 24483375..24484666 | 292     | 292      | RP5-1065J22.4 | RP5-1065J22.4 | 0        | 30491    | 1.34   | Heterologous Nuclear Ribonucleoprotein U (transcripts 1, 3, 4, 8) |
| 2             | 3271221..32716338 | 318     | 318      | LINCO0486 | AL121655.5 | 10988   | 17        | 0.02   | Long Intraspecific Non-Protein Coding RNA 468 |
| 2             | 47335292..47335571 | 280     | 280      | CALM2   | ICK      | 158690  | AC073283.4 | 0.78   | Autophagy-Related 16-like 1 / Small Cajal-body RNA 5 |
| 2             | 58190211..58190485 | 274     | 274      | SCARNA5 | AC017104.2 | 14386   | CERK1     | 0.54   | Kinesin family member 5C (transcript KIF5C_2) |
| 2             | 23145067..23146688 | 282     | 282      | RP11-119M12.4 | RP11-119M12.4 | 0        | 5099     | 0.78   | Long Intraspecific Non-Protein Coding RNA 468 |
| 2             | 26204664..26204971 | 278     | 278      | ACTB−119C2.1 | CTB−119C2.1 | 25518   | HNRNPA2B1 | 1.70   | Histone 1 (AD-agg. enriched) |
| 2             | 9903986..9904242 | 257     | 257      | MIR597  | LINC00599 | 162217  | LINC00599 | 1.50   | Long Intraspecific Non-Protein Coding RNA 468 |
| 2             | 5529123..5529444 | 322     | 322      | MIR59   | MIR59    | 33205   | ACTB     | 0.70   | Long Intraspecific Non-Protein Coding RNA 468 |
| 3             | 10179951..10179993 | 335     | 335      | MEG5/OGA | MEG5/OGA | 0       | 683       | 1.38   | Long Intraspecific Non-Protein Coding RNA 468 |

(Continues)
## TABLE 2 (Continued)

| Chrom. | Region            | Length | 5' gene | 5' dist. | 3' gene | 3' dist. | AD reads | AMC reads | AD/AMC reads | Chisq  | p <   | Transcript or encoded protein                                                                 |
|--------|-------------------|--------|---------|----------|---------|----------|----------|-----------|-------------|--------|--------|------------------------------------------------------------------------------------------------|
| 11     | 3563281..35663611| 331    | TRIM44  | 0        | KRT18P14| 196637   | 1147     | 592       | 1.94        | 0.0001 |        | Tripartite Motif Containing 44                                                                 |
| 11     | 62841562..62841839| 278    | RP11-727F15.9 | 7518 | WDR74  | 0        | 304      | 149       | 2.04        | 0.0001 |        | WD Repeat Domain 74                                                                               |
| 11     | 65502716..65503064| 349    | AP000769.7 | 4310 | MALAT1_1| 0        | 30539    | 16494     | 1.85        | 0.0001 |        | Metastasis Assoc. Lung Adenocarcinoma Transcr. 1 (Non-Protein)                                    |
| 11     | 111911504..111911780| 277    | RPL37AP8| 22029   | CRYAB   | 0        | 228      | 216       | 1.06        | N. S.  |        | Crystallin A beta, a small HSP (AD-agg-enriched)                                                  |
| 11     | 123061064..123061339| 276    | RPL31P47| 9662    | HSPA8   | 0        | 871      | 1512      | 0.58        | 0.0001 |        | Heat Shock Protein Family A (Hsp70) Member 8 (= HSC70)                                           |
| 12     | 110281619..110281909| 291    | ATP2A2  | 0       | RN75L769P| 78542    | 433      | 225       | 1.92        | 0.0001 |        | SERCA2 ATPase Sarcoplasmic/Endopl. Reticulum Ca2+ Transport2                                    |
| 13     | 45975357..45975649| 293    | AL445232.1 | 60203  | ZC3H13  | 0        | 372      | 240       | 1.55        | 0.0001 |        | Zinc Finger CCCH-Type Containing 13                                                              |
| 14     | 20343094..20343370| 277    | SNORD126| 16567   | RPPH1   | 0        | 2313     | 634       | 3.65        | 0.0001 |        | Ribonuclease P RNA Component H1                                                                  |
| 14     | 23321579..23321860| 282    | PABPN1  | 0       | SLC22A17| 24445    | 341      | 261       | 1.31        | 0.01   |        | Poly(A) Binding Protein Nuclear 1                                                                 |
| 14     | 49586595..49586871| 277    | RNA5SP384| 33846  | RPS29   | 0        | 22800    | 4958      | 4.60        | 0.0001 |        | Ribosomal protein, small subunit 29                                                             |
| 14     | 49853648..49853923| 276    | RNU6-539P| 13820   | Metazoa_ | 0        | 24010    | 7554      | 3.18        | 0.0001 |        | RNA, cytoplasmic Signal Recognition Particle                                                      |
| 14     | 102084750..102085033| 284    | RN75L472P| 7277    | HSP90AA1| 0        | 186      | 94        | 1.98        | 0.0001 |        | HSP 90 family (AD-agg-enriched)                                                                 |
| 14     | 6452910..6453209 | 300    | FAM64A  | 1440    | PITPNM3 | 0        | 671      | 520       | 1.29        | 0.004  |        | PITPNM3, Phosphatidylinositol Transfer Protein Memb.-Assoc.                                      |
| 15     | 1918769..19188246| 278    | KYNUP1  | 18300   | SNORD3A | 0        | 342      | 99        | 3.45        | 0.0001 |        | small nuclear RNA                                                                                 |
| 15     | 44911190..44911459| 270    | AC015936.3| 6643    | GFAP    | 0        | 7051     | 8989      | 0.78        | 0.0002 |        | Gial Fibrillary Acidic Protein, GFAP (AD-agg-enriched)                                           |
| 15     | 49814124..49814400| 277    | ACA2    | 163     | RP11-886H22.1 | 0  | 295  | 139  | 2.12        | 0.0001 |        | Transcribed processed pseudogene                                                                |
| 19     | 13298568..13298866| 299    | CTC-25014.1 | 139779  | CACNA1A | 0        | 10268    | 5984      | 1.72        | 0.0001 |        | Calcium Voltage-Gated Channel Subunit Alpha1 A                                                   |
| 19     | 36304598..36304871| 274    | CTD-316210.1 | 0     | LINC00665| 8195     | 556      | 164       | 3.39        | 0.0001 |        | Uncharacterized locus                                                                            |
| 19     | 44907737..44908015| 279    | APOE    | 0       | CTB-129P6.7 | 1359 | 1028 | 1003 | 1.02        | N. S.  |        | ApoE, Apolipoprotein E (AD-agg-enriched)                                                        |
| 19     | 48645763..48646036| 274    | DBP     | 8324    | CA11    | 0        | 407      | 573       | 0.71        | 0.0002 |        | Carbonic Anhydrase 11                                                                              |
| 19     | 49107777..49108115| 339    | SNRNP70 | 0       | LIN7B   | 6208     | 2096     | 1398      | 1.50        | 0.0001 |        | Small Nuclear Ribonucleoprotein U1 Subunit 70                                                    |
| 20     | 23637606..23637888| 283    | CST9    | 31729   | CST3    | 0        | 573      | 742       | 0.77        | 0.003  |        | Cystatin C, Cystatin 3                                                                            |
TABLE 2

| Chrom. Region | Length (nt) | 5' gene | 5' dist. | 3' gene | 3' dist. | AD reads | AMC reads | Chi² p | Transc. or encoded protein |
|---------------|-------------|---------|---------|---------|---------|----------|-----------|-------|---------------------------|
| 20 | 41533110..41533383 | 274 | RP4-620E11.5 | 150002 | 1.44 | 274 | RP4-620E11.5 | 150002 | 0.0001 | CHD6 Chromodomain Helicase DNA Binding Protein 6, CHDBP6 |
| 21 | 8258828..8259116 | 289 | RNA5-8S5 | 1894 | 1.44 | 274 | RNA5-8S5 | 1894 | 0.0001 | CHD6 Chromodomain Helicase DNA Binding Protein 6, CHDBP6 |
| 21 | 8495717..8495981 | 214 | APP | 1200 | 1.44 | 274 | APP | 1200 | 0.0001 | CHD6 Chromodomain Helicase DNA Binding Protein 6, CHDBP6 |
| 21 | 26021821..26022101 | 281 | LLPHP2 | 258487 | 1.44 | 274 | LLPHP2 | 258487 | 0.0001 | CHD6 Chromodomain Helicase DNA Binding Protein 6, CHDBP6 |
| 21 | 140784311..140784609 | 299 | LINC00632 | X | 11631 | 1200 | CDR1 | 11631 | 0.0001 | CHD6 Chromodomain Helicase DNA Binding Protein 6, CHDBP6 |

Note: All peaks shown were significant at p < 0.01; peak-coincident genes (at zero distance from peaks) are indicated by bold font. Red reads (AD/AMC) and p values for AD-AMC differences were not corrected for 1.6-fold higher RNA recovery from AD relative to AMC hippocampus, since all normalized RNA-seq libraries used 1 µg RNA. AD-agg enriched indicates proteins that were also found to be significantly enriched in AD aggregates relative to AMC.

2.6 Viral RNA and DNA fragments are enriched in AD aggregates relative to AMC

RNA and DNA fragments recovered from aggregates include sequences that do not map to the consensus human genome, but are related to known viral sequences compiled in the VirTect Database (Khan et al., 2019; Xia et al., 2019). After removal of all sequence reads homologous to the human genome, the remainder were mapped to the VirTect virus-sequence library. Raw viral RNA reads comprise 0.09% of total AMC RNA-fragment sequences (124,803/132,252,624), and 0.15% of AD RNA-fragments (236,939/158,235,930). Viral DNA reads comprise 0.33% of total AMC or AD DNA-fragment sequences (124,805/37,225,119 for AMC, 146,271/44,743,221 for AD). In view of their scarcity, such fragments are unlikely to drive aggregation; moreover, only a small minority of total raw reads met all three of the stringent VirTect thresholds (coverage depth ≥5x, a continuous/contiguous region ≥100 nt, and a read count ≥400) required for positive identification of human viruses. A total of 7 human viruses met all criteria, for a total of >135,000 reads (Table 4), out of >800 viruses or viral fragments detected (271,000 DNA reads, 362,000 RNA reads). As a negative control, the C. elegans genome was screened with identical parameters, yielding zero viral reads.

For 5 of the 10 viruses shown in Table 4, viral RNA fragments were significantly enriched in AD over AMC (at p < 0.01 to p < 0.0001), relative to the 1.6-fold AD enrichment of aggregate proteins (Ayyadevar, Balasubramaniam, Parcon, et al., 2016), for a combined significance of p < 1E-18. Three of these viruses, all linear duplex DNA viruses, were substantially more abundant in RNA than in DNA: Herpesvirus 2 (RNA/DNA = 9.1 in AMC, 17.7 in AD); Human Adenovirus 54 (RNA/DNA = 19.5 in AMC, 44.0 in AD); and Human Papillomavirus 72 (RNA/DNA = 15.2 in AMC, 21.0 in AD). Most of the remaining viruses were more highly represented in DNA than RNA, signifying that they were transcriptionally inactive. The consistently greater RNA/DNA ratios in AD tissue than in AMC is intrinsically "corrected" for relative viral and aggregate abundance, and strongly implies greater transcription and/or aggregation of RNA in AD hippocampus relative to AMC.
| A. Genes from E3, E4 reads (55 DAVID IDs, implicating 16 clusters of terms sharing members) |
|-------------------------------------|
| Cluster #, GO Term | Cluster Enrich. | Count | Fold Enrich. | Benjamini |
| 1. Extracellular exosome | 10.45 | 37 | 5.5 | 9E-21 |
| Acetylation | 10.45 | 31 | 4.2 | 3E-12 |
| Phosphoprotein | 10.45 | 35 | 2.0 | 8E-6 |
| Nucleus | 10.45 | 27 | 2.1 | 4E-4 |
| 2. Acetylation | 8.65 | 31 | 4.2 | 3E-12 |
| Poly(A) RNA binding | 8.65 | 21 | 7.1 | 1E-10 |
| Ubiquitinlike (Ubl) conjugation | 8.65 | 20 | 5.5 | 3E-8 |
| 3. Ubiquitinlike (Ubl) conjugation | 6.79 | 20 | 5.5 | 3E-8 |
| Cadherin binding, cell-cell adhesion | 6.79 | 9 | 12 | 6E-5 |
| Cell-cell adherens junction | 6.79 | 8 | 12 | 2E-5 |
| 4. Glycoprotein binding | 3.58 | 5 | 30 | 6E-4 |
| Protein stabilization | 3.58 | 6 | 17 | 4E-3 |
| 5. Myelin sheath | 2.77 | 8 | 22 | 2E-6 |
| Response to unfolded protein | 2.77 | 5 | 45 | 2E-3 |
| Protein refolding | 2.77 | 4 | 100 | 2E-3 |
| Stress response | 2.77 | 5 | 23 | 1E-3 |
| Chaperone | 2.77 | 6 | 14 | 1E-3 |

| B. Genes from AD, AMC reads (38 DAVID IDs, implicating 9 clusters of terms sharing members) |
|-------------------------------------|
| Cluster #, GO Term | Cluster Enrich. | Count | Fold Enrich. | Benjamini |
| 1. Intracellular ribonucleoprotein complex | 3.35 | 6 | 26 | 4E-4 |
| Methylation | 3.35 | 10 | 6.4 | 4E-4 |
| Ubiquitinlike (Ubl) conjugation | 3.35 | 10 | 3.8 | 2E-2 |
| Poly(A) RNA binding | 3.35 | 9 | 4.2 | 5E-2 |
| Acetylation | 3.35 | 12 | 2.3 | 8E-2 |
| 2. Extracellular matrix | 2.92 | 6 | 12 | 4E-3 |
| Chaperone | 2.92 | 4 | 13 | 4E-2 |
| ATPase activity | 2.92 | 4 | 12 | 0.13 |
| 3. Nucleoplasm | 2.82 | 15 | 3.2 | 9E-6 |
| Nucleus | 2.82 | 16 | 2.0 | 4E-5 |
| 4. Myelin sheath | 2.17 | 5 | 19 | 5E-3 |
| Chaperone | 2.17 | 4 | 13 | 4E-2 |
| Unfolded protein binding | 2.17 | 3 | 14 | 0.30 |

| C. Annotations from all RNA reads combined |
|-------------------------------------|
| GO Term | p value | Count | Fold Enrich. | Benjamini |
| Extracellular exosome | 2E-17 | 42 | 4.0 | 3E-15 |
| Extracellular matrix | 4E-16 | 18 | 16.3 | 4E-14 |
| Poly(A) RNA binding | 4E-13 | 26 | 5.7 | 8E-11 |
| Phosphoprotein | 4E-11 | 55 | 2.0 | 2E-09 |
| Acetylation | 2E-11 | 36 | 3.2 | 2E-09 |
| Isopeptide bond | 3E-11 | 22 | 5.9 | 2E-09 |
| Ubiquitinlike (Ubl) conjugation | 4E-11 | 26 | 4.6 | 2E-09 |
| Focal Adhesion | 2E-09 | 14 | 9.6 | 1E-07 |
| Myelin sheath | 4E-09 | 10 | 17.6 | 2E-07 |

(Continues)
2.7 | Cotranslational aggregation

As noted above, RNA reads substantially exceeded DNA reads by twofold to fivefold (Figure 2A). The propensity for nucleic acid-binding proteins to be inherently disordered, suggested above as an explanation for entrapment of nucleic acids in aggregates, is not expected to differ greatly between RNA- and DNA-binding proteins. We propose another mechanism, specific to RNA, that would account for the greater abundance of RNA in aggregates: cotranslational misfolding. Among the RNAs identified in AD-model aggregates (Table 2), many encode proteins that are themselves enriched in AD aggregates: for example, HnRNP_A2/B1, clusterin/ApoJ, β-crystallin A (CRYAB), SERCA_2/ATP2A2, GFAP, APOE, and Amyloid Precursor Protein/APP (Ayyadevara, Balasubramaniam, Parcon, et al., 2016). Of the 49 genes with RNA positively identified in aggregates, 9 (18%) encoded proteins that were also identified in aggregates. Twenty-three (46%) of the same 49 RNAs were significantly more abundant in AD aggregates than in controls, while 6 (12%) were significantly enriched in AD aggregates as both RNA and protein.

During translation, nascent proteins are at high risk for misfolding and aggregation until entire structural domains have emerged from the ribosome. From bacteria to mammals, chaperone complexes that include members of the HSP40, HSP60, and HSP90 families are closely associated with ribosomes, where they counteract misfolding of nascent polypeptides (Deuerling et al., 2019; Zhang & Ignatova, 2011). Nevertheless, the fraction of newly synthesized proteins that is cotranslationally degraded can exceed 50% (Turner & Varshavsky, 2000), indicating that chaperone protection is highly fallible. We wondered whether the remarkable abundance in aggregates of diverse RNA fragments, the great majority of which contain coding sequences, might be a clue that cotranslational aggregation occurs when misfolded, nascent proteins are neither prevented from misfolding nor degraded, prior to their coalescence with other misfolded proteins to form insoluble aggregates.

If this is the case, then interventions that arrest or delay translation should sharply reduce the aggregate content of RNA fragments. We used shRNA knockdown of EEF2 mRNA, reducing its steady-state level by 33% (Figure 3A,B) to attenuate protein translation in SY5Y-APPsw human neuroblastoma cells. Suppression of EEF2 has been shown to extend lifespan, reduce stress response, and improve the balance of protein quality control (Anisimova et al., 2018; David et al., 2010; Tavernarakis, 2008; Turner & Varshavsky, 2000). While prior research showed the existence of co-translational protein misfolding and degradation (G. Zhang & Ignatova, 2011), our results suggest that slowing translation may reduce aggregation of misfolded proteins, both in C. elegans (data not shown) and in cultured human

| GO Term                                      | p value  | Count | Fold Enrich. | Benjamini |
|----------------------------------------------|----------|-------|--------------|-----------|
| Membrane                                    | 2E-08    | 27    | 3.3          | 1E-06     |
| Identical protein binding                   | 3E-08    | 17    | 5.6          | 3E-06     |
| Intracellular ribonucleoprotein complex     | 4E-08    | 9     | 17.7         | 1E-06     |
| Disease mutation                            | 4E-08    | 27    | 3.2          | 1E-06     |
| Methylation                                 | 1E-07    | 17    | 5.1          | 3E-06     |
| Cytosol                                     | 1E-07    | 32    | 2.6          | 4E-06     |
| Glycoprotein binding                        | 2E-07    | 7     | 26.7         | 2E-05     |
| Protein binding                             | 1E-06    | 55    | 1.6          | 6E-05     |
| Oxidation                                   | 2E-06    | 6     | 56           | 2E-05     |
| Chaperone                                   | 4E-06    | 8     | 12           | 9E-05     |
| Unfolded protein binding                    | 5E-06    | 7     | 15.8         | 2E-04     |
| Neurodegeneration                            | 5E-06    | 9     | 9.3          | 1E-04     |
| Protein stabilization                       | 1E-05    | 7     | 13.5         | 9E-03     |
| Response to unfolded protein                | 2E-05    | 5     | 31.2         | 7E-03     |
| Cadherin binding, cell-cell adhesion        | 2E-05    | 9     | 7.7          | 7E-04     |
| Protein refolding                           | 2E-05    | 4     | 70           | 6E-03     |
| Cell-cell adherens junction                 | 2E-05    | 9     | 7.5          | 5E-04     |
| MHC class II protein complex binding        | 3E-05    | 4     | 62.1         | 1E-03     |
| Amyloidosis                                 | 1E-04    | 4     | 43.2         | 2E-03     |
| Stress response                             | 3E-04    | 6     | 15           | 5E-03     |

Note: N.B.: Minor terms were omitted from each cluster. Cluster enrichment is the “Enrichment Score” from Functional Annotation Clustering under DAVID; fold enrichment is “Fold Change” per term; Benjamini indicates the false discovery rate, FDR, predicted by the Benjamini-Hochberg procedure.
cells as follows. In SYSY-APPsw cells, shRNA targeting EEF2 eliminated over 90% of the RNA entrapped in aggregates (p < 0.0001; Figure 3C,D), far exceeding the 33% efficacy of EEF2 knockdown (Figure 3B). At the same time, this RNAi exposure had little or no effect on aggregate DNA content (Figure 3E,F), but reduced aggregate protein by 30% (p < 0.01; Figure 3G,H). In SYSY-APPsw cells treated for 4 h with MG132, a cell-permeant proteasome inhibitor, aggregates increased 20–30%; however, this rise was not accompanied by any increase in aggregate RNA fragments (Figure S3). This suggests that the reduction in aggregate burden per se cannot account for the decline in aggregate RNA after EEF2 knockdown.

3 | DISCUSSION

Pathognomonic complexes associated with neurodegenerative diseases, including Alzheimer’s, Parkinson’s, and Huntington’s diseases, are widely termed “protein aggregates” because their diagnostic antigenic markers are proteins. Whether these aggregates also contain other components, however, is a question that has not been adequately addressed. We were aware that some amalgamations of cell debris that accumulate with aging, known as lipofuscin granules, contain a complex mixture of oxidized, glycated and carbonylated proteins, lipids, and possibly other carbohydrates; however, nucleic acids were only rarely noted among their constituents (Cindrova-Davies et al., 2018; Nowotny et al., 2014). Ginsberg et al. (1998, 1999) reported that 80% of neurofibrillary tangles and 55% of senile (amyloid) plaques can be stained with acridine orange, implying the presence of RNA. Numerous studies have implicated nucleic acid binding by mammalian prion-like protein, PrP (Cordeiro et al., 2014; Gomes et al., 2012; Macedo et al., 2012; Silva et al., 2008), and the evidence that this extends to other neurodegenerative-disease seed proteins has been reviewed (Cordeiro et al., 2014).

We were led from the results of proteomic “contactome” studies, intended to define the molecular architecture of aggregates (i.e., which proteins adhere to which other proteins), to investigate the nature, extent, and specificity of nucleic acids incorporated into aggregates. In each of these three respects, the results were unexpected. We observed two- to fivefold more RNA than DNA in aggregates, whether isolated from AD or control hippocampus (Figure 2A). Many RNA sequences identified in human hippocampal aggregates were differentially abundant in AD- vs. control-derived aggregates; of these, twice as many were enriched significantly in AD aggregates as in non-AD controls. Proteomic analyses of aggregates from equal-weight aliquots of AD vs. AMC hippocampus samples indicate an AD/AMC ratio of 1.84 (t test p < 0.01), in reasonable agreement with previous AD/AMC protein ratios of 1.65 and 1.66 for Aβ and tau aggregates, respectively (Ayyadevara, Balasubramaniam, Parcon, et al., 2016), and do not differ significantly from the ratios observed here for RNA and DNA. Interactome complexities of Aβ and tau aggregates (unpublished data) indicate AD/AMC ratios of 1.84 and 1.64, respectively (each t test p < 0.01)—implying more abundant and varied protein interfaces in AD than in AMC hippocampus.
When we compared aggregates isolated from glioblastoma cells overexpressing an APOE3 vs. APOE4 transgene, sequences with the most differential representation were quite consistently more abundant in APOE3-bearing cells. We believe this very likely reflects the surprising ability of APOE4 to enter nuclei and bind competitively to the CLEAR/E-box motifs recognized by transcription factor EB (TFEB), thereby inhibiting expression of autophagy and lysosomal genes (Parcon et al., 2018). Not surprisingly, >90% of aggregate DNA originated from nuclear aggregates, while RNA in aggregates was predominantly of cytoplasmic origin.

Only a small fraction of aggregate-associated nucleic acids (0.09 – 0.15% of RNA reads, 0.33% of a smaller set of DNA reads) appears to be of viral origin, although these totals may be underestimated due to as-yet-uncatalogued and mutated viruses or endogenous retroposons (Sanjuan et al., 2010). The striking 2.3-fold enrichment of viral RNA sequences in AD aggregates relative to controls, vs. only 1.15-fold for viral DNA fragments (see Table 4), is consistent with possible roles of viral infection and/or transcriptional activation in the etiology of Alzheimer’s disease (Balin & Hudson, 2018; Irish et al., 2009; Kreutz, 2002; Kristensson, 1992; Linet et al., 1997; Romeo et al., 2019; Steel & Eslick, 2015). It is also possible that the observed enrichments reflect secondary effects of Alzheimer’s pathology, including chronic low-grade inflammation (Majde, 2010), insofar as they may augment viral infection or transcriptional activation in the AD brain. Previous studies have shown that soluble amyloid-like proteins bind to nucleic acids, which could lead to formation of amyloid fibrils (Di Domenico et al., 2012). Nucleic acid-containing amyloid fibrils induce interferon and activate innate immune Toll-like receptors, driving neuroinflammation and synapse loss in AD (Di Domenico et al., 2012; Roy et al., 2020).

Somaically integrated and even endogenous (heritable) viral genomes have highly variable insertion sites. As a result, viral RNAs
and DNAs require identification by searching a database of known virus genomes. Quantitation of viral sequences may thus be underestimated due to the many human viruses as yet unidentified, plus the high viral mutation rates impeding sequence alignment. Nevertheless, viral RNA and DNA comprise very small fractions of the nucleic acids recovered from aggregates. From an evolutionary perspective, however, they may ultimately be responsible for the perseverance in our genomes of proteins with high levels of disorder and high probability of aggregation—provided only that disorder contributes to the ability to bind viral nucleic acids and/or to sequence them in aggregates.

The observed data are consistent with a scenario in which endogenous retroviruses—of which HERV K113 is the youngest and only actively transposing exemplar (Boller et al., 2008)—and integrated genomic copies of retroviruses (e.g., Hepatitis C viruses) and DNA viruses (e.g., Herpes viruses) become activated and transcribed into RNA. Darwinian selection might favor protein variants that are predisposed to misfold, provided that they disable replication and transcription of viral genomes within cells by entrapping them in aggregates. Variants that enhanced survival of a pandemic by even a few percent would undergo strong selective pressure to sweep the population, becoming the predominant or sole alleles (Karlsson et al., 2014).

Predicted G-quadruplex-forming sequences in both DNA and RNA, the best known and most abundant class of four-stranded nucleic acid structures, are also markedly enriched in AD aggregates. Sequences with G-quadruplex-forming potential can be recognized by their binding proteins based on singular structural features; they thus often serve as recognition sites for critical proteins with key surveillance or regulatory functions, such as telomere-binding proteins, viral-replication proteins, and gene promoter regions (Brazda et al., 2014).

The observation of consistent functional-annotation terms and clusters, both within each aggregate type and between the two sources of aggregates, confirms that the particular RNA species found in aggregates are not a random sampling from the transcriptome—but it does not explain the basis for their enrichment. We propose two routes by which nucleic acids can be incorporated into aggregates that form either as a result of aging per se or due to an age-dependent pathology such as Alzheimer’s disease: (1) “hitchhiker” or “bystander” entrapment of DNA and RNA, when they are bound by proteins that become misfolded and consequently enmeshed in aggregates; and (2) cotranslational misfolding of proteins in the midst of their translation, which might be expected to also ensnare ribosomes and the mRNAs they are translating. The first mechanism is supported by the remarkably high abundance of DNA- and RNA-binding proteins in the aggregate interactome (Figure 1).

The second mechanism is most compellingly supported by the decimation (>10-fold reduction) of aggregate RNA content following shRNA knockdown of the translational process factor EEF2. We suspect that cotranslational aggregation occurs preferentially in pathways or processes that involve enzymes with multiple partners, and/or several nucleic acid-binding proteins—thus accounting for the highly significant enrichments observed in aggregated RNA, for genes annotated with specific clusters of descriptive terms. Note that neither of these explanations attributes a primary or causal role to nucleic acids, through which they would "drive" aggregate accrual. Rather, they are collateral casualties due to misfolding of their attached proteins.

Why did EEF2 knockdown have a far greater effect on RNA content than protein content of aggregates? This is actually the expected result if cotranslational aggregation accounts for only a minor fraction of the protein deposition in aggregates, but is responsible for 90% of their RNA content. Nascent proteins may misfold transiently during translation, but even mature proteins can misfold over time, as a consequence of post-translational disturbances such as oxidation, phosphorylation, or alkylation, and other temperature- or time-dependent processes that favor misfolding of pre-existing proteins. Such processes would continue with little prospect of reversal, for all previously-synthesized proteins—unabated by translational arrest. RNA, however, may only appear in aggregates when it is bound by a misfolded (and hence aggregation-prone) protein, or when the RNA is in the process of translation into a nascent protein that has a high probability of transient misfolding and aggregation. Our observations imply that cotranslational aggregation is the predominant route, accounting for at least 90% of aggregate RNA.

Our data suggest that proteostasis in SYSY-APP cells, which are subjected to chronic ER stress by continual generation of Aβ1–42, is normally insufficient to prevent cotranslational aggregation. However, even moderate alleviation of that stress appears to shift the balance back to sustainable translational proteostasis. Translational inhibition has been reported to lower chronic inflammation (Mazumder et al., 2010), which may be a consequence of reduced protein aggregation, augmented by a disproportional decrease in aggregate RNA.

### 4 | CONCLUSIONS

"Protein aggregates" contain nucleic acid constituents that are highly nonrandom in sequence—making it unlikely that they are artifacts, but instead implying that they contain protein-binding features (including G-quadruplexes) that might pull them into aggregates. The number and variety of viral sequences found in aggregates suggests that there may be an evolutionary advantage (i.e., antiviral protection) to the synthesis of nucleic acid binding proteins that readily misfold and thus sequester viral genomes in aggregates. Significant enrichment of viral sequences in AD aggregates, relative to controls, is consistent with roles for integrated viruses in AD susceptibility. The preferential enrichment of RNA over DNA in aggregates may implicate a mechanism specific to transcripts: cotranslational aggregation of polysomes during initial misfolding of nascent polypeptides. This process would likely be quite sensitive to the balance between translation rate and chaperone-mediated refolding capacity. A critical role of cotranslational entrapment is supported by our observation that shRNA...
knockdown of the translation elongation factor EEF2, although only 25–35% effective, selectively eliminates at least 90% of RNA in aggregates.

5 | EXPERIMENTAL PROCEDURES

5.1 | Preparation of cells

Cells were grown in 75-cm² flasks, with culture medium comprising Dulbecco’s Modified Eagles Medium (DMEM) supplemented with 10% (v/v) fetal bovine serum (FBS) at 37°C, grown in an atmosphere of air supplemented with 5% CO₂. Cells were harvested and washed with phosphate buffered saline and then digested with 0.25% (w/v) trypsin (Thermo Fisher) at 37°C for 4 min or until cells detach from the surface.

5.2 | EEF-2 knockdown

For EEF-2 gene knockdown, RNAi knockdown was performed with 3 distinct EEF-2 shRNA sequences, targeting human EEF-2 (SASI_Hs01_00212218 and SASI_Hs_0022218 from Millipore-Sigma; s4493 from Thermo Fisher), each introduced separately into SH-SYSY-APP cells. Cells were harvested and replated at a density sufficient to achieve ~80% confluence 72 h later. RNAiMax (Thermo Fisher) was used as the transfection reagent, following the manufacturer’s protocol. MISSION shRNA universal negative controls (SICO01 and SICO02, Millipore-Sigma) were transfected by the same protocol, as negative controls for the EEF-2 knockdowns. Cells were harvested and flash frozen 72 hours after transfection.

5.3 | Isolation of sarkosyl-insoluble aggregates

Aggregates were prepared from Alzheimer’s Disease (AD) vs. age-matched control (AMC) hippocampus; T98G/APOE3 or T98G/APOE4 human glial cell pellets; or SY5Y-APPsw human neuroblastoma cell pellets. Frozen tissues or cells were pulverized with a mortar and pestle (cooled on dry ice) and suspended in lysis buffer containing 20-mM HEPES pH 7.4, 0.3-M NaCl, 2-mM MgCl₂, 1% NP40 (w/v), supplemented with phosphatase and protease inhibitors (CalBiochem). Tissue suspensions were lysed in a Teflon homogenizer (2 times 10 s, at 0°C) and sonicated (3 times 10 s, at 0°C). Samples were centrifuged 5 min at 600 × g to remove debris. Supernatant protein was quantified and each sample (0.6–1.0 mg) was centrifuged 15 min at 13,000 × g. Supernatants (soluble protein) were removed, and to each insoluble pellet the same lysis buffer was added plus 1% (v/v) sarkosyl, and mixed well. Samples were centrifuged 20 min at 100,000 × g; supernatants and pellets were recovered as “sarkosyl-soluble aggregates” and “sarkosyl-insoluble aggregates”, respectively.

5.4 | Immunoprecipitation of amyloid beta and tau aggregates

AD and AMC hippocampal tissue samples were pulverized as described above. After removal of debris (centrifugation for 5 min at 1400 × g), protein was quantified by the Bradford protein assay. Protein was then gently mixed with magnetic beads coated with antibody to either Aβ1-42 (ab11132) or tau (ab80579) for immunopulldown (IP); sarkosyl-insoluble protein was isolated from the antibody-bound fractions as described previously (Ayyadevara, Balasubramaniam, Parcon, et al., 2016).

5.5 | Aggregate contactome generation

Insoluble aggregates isolated from SY5Y-APPsw cells as above, were cross-linked following procedures described previously (Balasubramaniam et al., 2019). In brief, purified aggregates were rinsed, cross-linked with modified click reagents, digested with trypsin, and the linked peptide pairs were affinity purified using streptavidin-coated beads to capture the biotin-coupled crosslinking moiety. Cross-linked peptide pairs were identified from high-resolution LC/MS-MS raw data files, using a modified version of Xlink identifier (Balasubramaniam et al., 2019; Du et al., 2011). Xlink identifier outputs were analyzed with the GePhi software package to calculate the degree (number of interacting partners) of each hub. Because high-molecular-weight proteins (e.g., titin) have greater potential to interact with other proteins, spectral hits for each hub were normalized, i.e. divided by the length of that hub protein in amino acids. Identified contactome proteins were categorized by degree, as described previously (Balasubramaniam et al., 2019). Proteins with a high normalized degree (number of interacting partners divided by length in amino acids) or classified as hub-connectors (connecting 2 or more hubs that are not otherwise connected) were pursued by further graph modeling; the Cytoscape package (Shannon et al., 2003) was used with default parameters to construct and visualize graphs.

5.6 | Isolation and quantitation of nucleic acids in aggregates

For sequencing of nucleic acid fragments from isolated aggregates, RNA and DNA were extracted from sarkosyl-insoluble material isolated from cultured cells, or from AD and AMC hippocampus, using the Qiagen AllPrep kit following manufacturer’s instructions and a protocol in which this kit was shown to recover even small nucleic acid fragments (Pena-Llopis & Brugarolas, 2013).

To quantify DNA and RNA trapped in sarkosyl-insoluble aggregates, nucleic acids were extracted and assayed by multiple protocols, with consistent results. These consisted of (1) separation of RNA and DNA fragments using a Qiagen AllPrep DNA/RNA extraction kit according to the manufacturer’s protocol, with recovery
assayed by absorbance at 260 nm; (2) separation of RNA and DNA fragments with the Qiagen kit, and quantitation by ethidium bromide and/or SYBR Gold after resolution by acrylamide gel electrophoresis; (3) selective enzymatic digestion with RNase-free DNase (Thermo Fisher, CA) and assay by 260-nm absorption (RNA directly; DNA by subtraction), and (4) using TRI Reagent (Molecular Research Ctr., TR118) to isolate RNA, DNA, and protein in a single protocol. Figure 2 data were obtained by method (3) above.

5.7 | RNA-seq and ChIP-seq analyses

All RNA-seq and ChIP-seq analyses were performed by the UT Southwestern Genomics Core, analyzed using the CLC Genomics Workbench. We employed ChIP-seq to evaluate DNA-fragment specificity; thus, the primary analytic value is the number of significant peaks, with peak validity assessed by an E value relative to a flat distribution (peak absence). RNA-seq was preceded by peak validation, just as for ChIP-seq. Subsequently, valid-peak reads that map uniquely to exons (“unique exon reads”) were summed as our expression metric, and were used to determine differential expression between groups.

The nucleic acid contig assemblies were quite consistent in size, 579 ± 34 (SD) base pairs in length for DNA peaks, and 291 ± 31 (SD) for RNA-fragment contigs. The efficiency of ChIP-seq and RNA-seq fragment cloning protocols, employed prior to sequencing, is quite sensitive to fragment size. Under normal ChIP-seq protocols, they would be determined by shearing or sonication, size selection by cloning vector, and/or manual size selection. However, in the case of aggregate nucleic acid fragments, other factors may be influential—such as the size, age, and intracellular location of individual aggregates.

5.8 | Viral sequence analysis

We employed a modified version of VirTect to scan DNA and RNA fragment sequences from human AD and AMC (age-matched control) hippocampi. VirTect is a pipeline script that calls a sequence of RNA-seq pattern-matching routines (Khan et al., 2019). VirTect retrievals of viral matches to aggregate nucleic acid reads, from 3 AD and 3 AMC brain samples, were filtered using the following parameters (https://github.com/WGLab/VirTect/blob/master/README.md): ≥5x coverage depth, a continuous/contiguous region cutoff of ≥75, and a read count ≥50. Several protocol modifications were made for our pipeline: (1) tophat2 was replaced by hisat2; (2) code was optimized to use all available threads; (3) the internal threshold number of reads was reduced in exploratory runs for the purpose of obtaining AD/AMC read ratios, but recommended thresholds were maintained to eliminate false positives in the assignment of valid hits; (4) the modified script was rewritten in Bash, with unnecessary subroutine calls deleted to reduce run-time. The database screened by VirTect comprises complete sequences of 757 viruses, as described (Khan et al. 2019). Target sequences were not restricted to human viruses, in recognition of the high frequency of zoonoses and multiple-host pathogens.

5.9 | G-quadruplex analyses

We employed two programs to screen RNA and DNA sequences for G-quadruplex-forming regions: G4CatchAll (Dolca, 2019) and QGRS Mapper (Kikin et al., 2006). Both strands were scanned for each DNA-fragment sequence, but only strands with G4-forming potential were pursued in subsequent analyses. The following parameters were used for G4CatchAll: G3L (loop limit) was set to 1.3; G2L (allowing 2-G loops) was set to 1.3; G4H (enables the G4Hunter algorithm for final evaluation). The following parameters were used for QGRS Mapper: Max. Length 30; Min G-Group 2; Loop size 0 – 36.

5.10 | Statistical analyses

Inter-group differences were tested for significance by 2-tailed Behrens–Fisher heteroscedastic t tests, unless otherwise indicated. These conservative tests are appropriate to small–sample comparisons in which the intra-group variance is not well estimated. Comparisons of ratios generally employed Yates chi² (chi-squared) nondirectional tests, substituting 2-tailed Fisher Exact tests as required to meet numerical constraints. This conservative replacement is stated in the text but is not made explicit (line by line) in tables to conserve space.

ACKNOWLEDGMENTS

This work was supported by grants (Merit 2 I01 BX001655 and Senior Research Career Scientist Award IK6 BX004851) to R. J. S. R. from the U. S. Dept. of Veteran Affairs; by R01 grant AG062254–01 to R. J. S. R and S. A., and by Program Project grant 2P01AG012411-17A1 (W. S. T. Griffin, P. I.), both from the National Institute on Aging (NIA/NIH).

CONFLICTS OF INTEREST

The authors declare no competing or conflicting interests.

AUTHOR CONTRIBUTIONS

RJSR and SA designed the study; MB and SA performed aggregate cross-linking studies; MB undertook subsequent data analysis and network modeling; MB and AG performed additional proteomics data analyses and contactome construction; JJ undertook G-quadruplex and human-virus searches, and checked all RNA and DNA sequencing data analyses (based on data and initial CLC bio/Qiagen analyses provided by the Genomics Core facility, University of Texas Southwest Medical Center); RA and SA performed EEF2-shRNA knockdown experiments, and RNA and DNA recovery quantitations; RJSR wrote the manuscript with contributions from SA and MB. All authors read and approved the final manuscript.
All data generated or analyzed during this study are included in this published article and its supplementary information files. Any reasonable request for additional data will be honored.
in Alzheimer disease. *Journal of Clinical Investigation*, 130(4), 1912–1930. https://doi.org/10.1172/JCI133737

Sanjuan, R., Nebot, M. R., Chirico, N., Mansky, L. M., & Belshaw, R. (2010). Viral mutation rates. *Journal of Virology*, 84(19), 9733–9748. https://doi.org/10.1128/JVI.00694-10

Sardiello, M. (2016). Transcription factor EB: from master coordinator of lysosomal pathways to candidate therapeutic target in degenerative storage diseases. *Annals of the New York Academy of Sciences*, 1371(1), 3–14. https://doi.org/10.1111/nyas.13131

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498–2504. https://doi.org/10.1101/gr.1239303

Shi, Y., & Berg, J. M. (1995). Specific DNA-RNA hybrid binding by zinc-finger proteins. *Science*, 268(5208), 282–284. https://doi.org/10.1126/science.7536342

Silva, J. L., Lima, L. M., Foguel, D., & Cordeiro, Y. (2008). Intriguing nucleic-acid-binding features of mammalian prion protein. *Trends in Biochemical Sciences*, 33(3), 132–140. https://doi.org/10.1016/j.tibs.2007.11.003

Steel, A. J., & Eslick, G. D. (2015). Herpes viruses increase the risk of Alzheimer’s disease: A meta-analysis. *Journal of Alzheimer’s Disease*, 47(2), 351–364. https://doi.org/10.3233/JAD-140822

Stoyanova, T., Roy, N., Kopañda, J., Raychaudhuri, P., & Bagchi, S. (2009). DDB2 (damaged-DNA binding protein 2) in nucleotide excision repair and DNA damage response. *Cell Cycle*, 8(24), 4067–4071. https://doi.org/10.4161/cc.8.24.10109

Tavernarakis, N. (2008). Ageing and the regulation of protein synthesis: a balancing act? *Trends in Cell Biology*, 18(5), 228–235. https://doi.org/10.1016/j.tcb.2008.02.004

Turner, G. C., & Varshavsky, A. (2000). Detecting and measuring cotranslational protein degradation in vivo. *Science*, 289(5487), 2117–2120. https://doi.org/10.1126/science.289.5487.2117

Turner, M., & Hodson, D. J. (2012). An emerging role of RNA-binding proteins as multifunctional regulators of lymphocyte development and function. *Advances in Immunology*, 115, 161–185. https://doi.org/10.1016/B978-0-12-394299-9.00006-0

Wade, P. A. (2001). Methyl CpG-binding proteins and transcriptional repression. *BioEssays*, 23(12), 1131–1137. https://doi.org/10.1002/bies.10008

Xia, Y., Liu, Y., Deng, M., & Xi, R. (2019). Detecting virus integration sites based on multiple related sequencing data by VirTect. *BMC Medical Genomics*, 12(Suppl 1), 19. https://doi.org/10.1186/s12920-018-0461-8

Zacco, E., Grana-Montes, R., Martin, S. R., de Groot, N. S., Alfano, C., Tartaglia, G. G., & Pastore, A. (2019). RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43. *Journal of Molecular Biology*, 431(8), 1671–1688. https://doi.org/10.1016/j.jmb.2019.01.028

Zhang, G., & Ignatova, Z. (2011). Folding at the birth of the nascent chain: coordinating translation with co-translational folding. *Current Opinion in Structural Biology*, 21(1), 25–31. https://doi.org/10.1016/j.sbi.2010.10.008

Zhang, T., Faraggi, E., Li, Z., & Zhou, Y. (2013). Intrinsically semi-disordered state and its role in induced folding and protein aggregation. *Cell Biochemistry and Biophysics*, 67(3), 1193–1205. https://doi.org/10.1007/s12013-013-9638-0

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Shmookler Reis RJ, Atluri R, Balasubramaniam M, Johnson J, Ganne A, Ayyadevara S.

“Protein aggregates” contain RNA and DNA, entrapped by misfolded proteins but largely rescued by slowing translational elongation. *Aging Cell*. 2021;20:e13326. https://doi.org/10.1111/acel.13326