nature communications

Article

https://doi.org/10.1038/s41467-024-52886-2

Transcript errors generate amyloid-like proteins in human cells

Received: 18 July 2023

Accepted: 23 September 2024

Published online: 07 October 2024

Check for updates

Claire S. Chung[1,11], Yi Kou[2,11], Sarah J. Shemtov[1,11], Bert M. Verheijen[1,11], Ilse Flores[3], Kayla Love[2], Ashley Del Dosso[4], Max A. Thorwald[1], Yuchen Liu[2], Daniel Hicks[1], Yingwo Sun[1], Renaldo G. Toney[1], Lucy Carrillo[1], Megan M. Nguyen[5], Huang Biao[4], Yuxin Jin[3], Ashley Michelle Jauregui[3], Juan Diaz Quiroz[6], Elizabeth Head[7], Darcie L. Moore[8], Stephen Simpson[9], Kelley W. Thomas[9], Marcelo P. Coba[3], Zhongwei Li[4], Bérénice A. Benayoun[1], Joshua J. C. Rosenthal[6], Scott R. Kennedy[5], Giorgia Quadrato[4], Jean-Francois Gout[10], Lin Chen[2] & Marc Vermulst[1] ✉

Aging is characterized by the accumulation of proteins that display amyloid-like behavior. However, the molecular mechanisms by which these proteins arise remain unclear. Here, we demonstrate that amyloid-like proteins are produced in a variety of human cell types, including stem cells, brain organoids and fully differentiated neurons by mistakes that occur in messenger RNA molecules. Some of these mistakes generate mutant proteins already known to cause disease, while others generate proteins that have not been observed before. Moreover, we show that these mistakes increase when cells are exposed to DNA damage, a major hallmark of human aging. When taken together, these experiments suggest a mechanistic link between the normal aging process and age-related diseases.

Protein aggregation is a defining hallmark of human aging and disease[1,2]. At a molecular level, protein aggregates are formed by misfolded proteins that form amorphous protein deposits or self-assemble into large, neatly organized amyloid fibers. These aggregates play an important role in various neurodegenerative diseases, including Alzheimer's disease (AD), Parkinson's disease (PD) and Creutzfeld–Jakob Disease (CJD)[3,4]. However, they also contribute to the functional decline associated with normal aging and the pathology of a variety of other age-related diseases, including cancer[5], amyotrophic lateral sclerosis (ALS), diabetes, heart disease and cataracts[6–9]. In familial cases of amyloid diseases, patients tend to carry a single point mutation that dramatically increases the amyloid propensity of the affected protein[10]. However, the majority of cases that enter the clinic are non-familial in nature, and it is currently unclear how these pathogenic proteins are generated in the absence of a well-defined genetic predisposition.

One long-standing hypothesis is that in non-familial cases of these diseases, amyloid proteins are generated by epi-mutations, non-genetic mutations that are only present in transcripts. For example, if a mistake was made during RNA synthesis[11–14] or RNA editing[15], a small cache of mutant proteins could be generated that displays amyloid or prion-like behavior. Although their initial number would be small, amyloid and prion-like proteins are defined by their ability to replicate themselves by binding to WT proteins through strong, non-covalent interactions and converting them to an amyloid state[16]. Through this self-templating mechanism, a small cache of mutant proteins created

[1]University of Southern California, Leonard Davis School of Gerontology, Los Angeles, USA. [2]University of Southern California, Molecular and Cellular Biology Department, Los Angeles, USA. [3]University of Southern California, Keck School of Medicine, Los Angeles, USA. [4]University of Southern California, Eli and Edythe Broad CIRM Center for Regenerative Medicine and Stem Cell Research, Los Angeles, USA. [5]University of Washington, Department of Pathology and Laboratory Medicine, Seattle, USA. [6]Marine Biological Laboratory, Bell Center, Woods Hole, USA. [7]University of California Irvine, Department of Pathology and Laboratory Medicine, Irvine, USA. [8]University of Wisconsin, Department of Neuroscience, Madison, USA. [9]University of New Hampshire, Department of Molecular, Cellular, & Biomedical Sciences, Durham, USA. [10]Mississippi State University, Department of Biology, Mississippi State, USA. [11]These authors contributed equally: Claire S. Chung, Yi Kou, Sarah J. Shemtov, Bert M. Verheijen. ✉e-mail: vermulst@usc.edu

by a transcript error could grow in size and number and eventually seed the amyloid fibers that characterize aging cells (Fig. 1A). Broadly speaking, these aggregates are divided into two categories: amyloid and amyloid-like structures. Amyloid structures are created by the self-assembly of amyloid proteins into highly ordered, fibrillar aggregates characterized by a cross-beta sheet structure. In contrast, amyloid-like structures consist of pathological protein aggregates that lack classic amyloid fibrils, but because the misfolded proteins that give rise to these structures share the characteristic self-templating behavior of amyloid proteins, they are often referred to as amyloid-like proteins.

Although a role for transcript errors in the formation of these structures has been suspected, it has proven difficult to test this hypothesis due to technical limitations that are related to error detection. To solve this problem, we recently optimized[17] an RNA sequencing tool termed circle-sequencing[18,19], which allows for high-fidelity sequencing of mRNA molecules (Supplementary Fig. 1). Here, we use circle-sequencing to demonstrate that transcript errors are ubiquitous in human cells, and that they indeed result in proteins with amyloid and prion-like properties. We support these observations with a variety of cellular, biochemical and biophysical experiments that demonstrate that the proteins generated by these errors can successfully convert WT proteins to an amyloid-like state, which then self-assemble into protein aggregates with a variety of structures. Finally, we show that the amount of mutant proteins required to initiate large-scale protein aggregation is routinely breached as a result of DNA damage, a ubiquitous hallmark of aging cells. As a result, our experiments establish a plausible, mechanistic link between DNA damage and protein aggregation, two major hallmarks of human aging and age-related diseases, including AD. In doing so, our observations provide fresh insight into the role of mutagenesis in human aging and disease, and suggest a plausible mechanism by which amyloid- and prion-like diseases can develop.

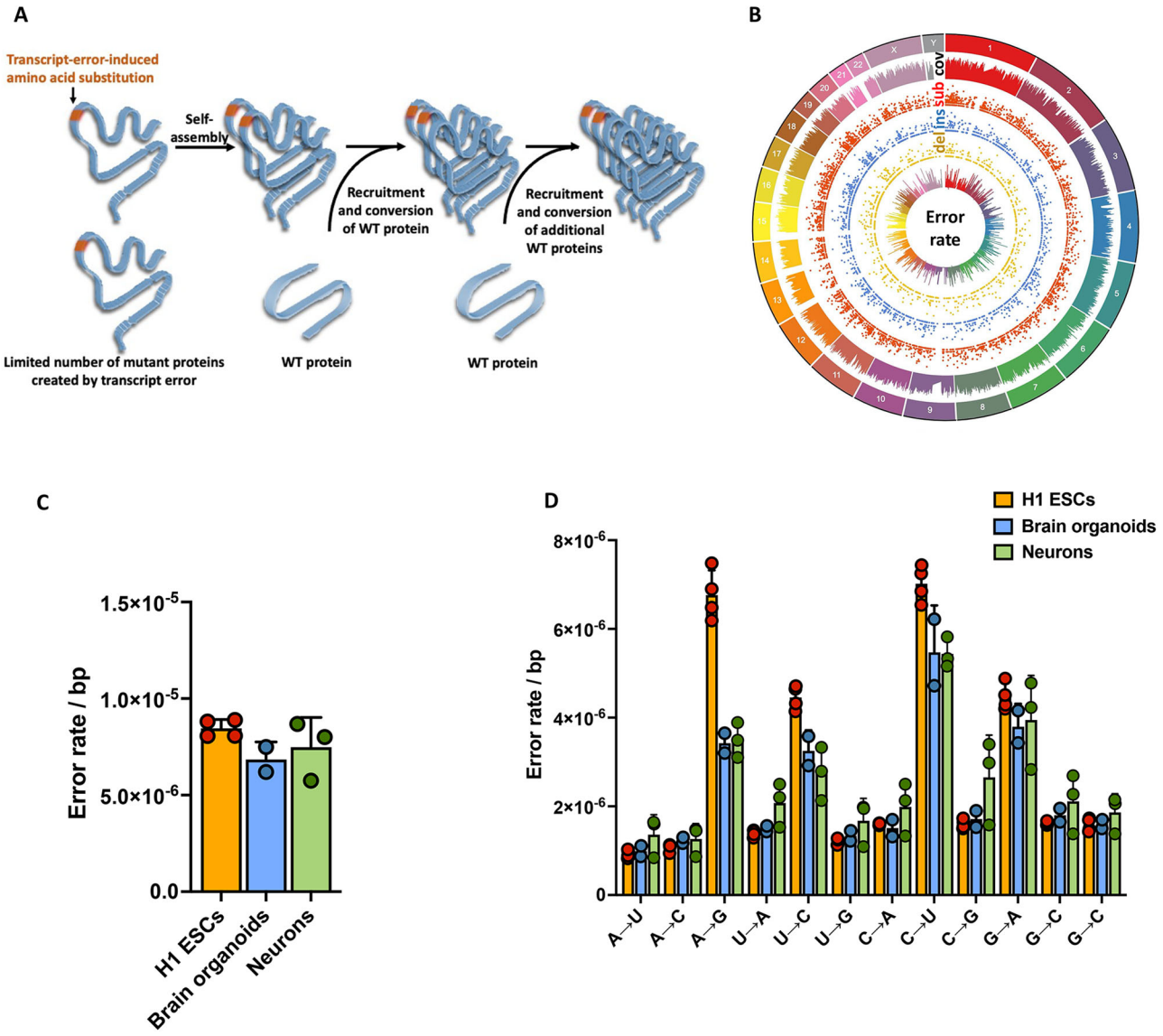


Fig. 1 | Graphical representation of hypothesis and summary of transcription error data. A We hypothesize that transcription errors could give rise to amyloid- and prion-like proteins. This relatively small cache of mutant proteins can then form a seed that recruits WT proteins and converts them to an amyloid-like state to generate the large amyloid fibers and amorphous deposits that characterize protein aggregation diseases. **B** Transcription errors were identified across the genome of H1 human embryonic stem cells ($n = 4$, ESCs), brain organoids ($n = 2$) and human neurons ($n = 3$). All replicates are independent biological replicates. **C**,**D** The error rate and spectrum of H1 ESCs, brain organoids and human neurons are similar, with the exception of A → G errors, which most likely indicate off-target A to I RNA editing (n is identical to **B**). Error bars indicate standard error of the mean. Source data are provided as a Source Data file.

Table 1 | Transcription errors affect proteins directly implicated in amyloid- and prion-like diseases

Gene	Protein	Disease	Errors detected	Mutations mimicked	Key aa's affected	Amyloid potential ⇧
ABri peptide	ITM2B	FBD and FDD	66	2	6	27
Amyloid precursor protein	APP	Alzheimer's disease	266	6	9	51
Cystatin-B	CSTB	EPM1	33	2	1	0
Fused in sarcoma	FUS	ALS	25	1	3	10
Gamma-crystallin D	CRYGD	Coralliform cataracts	1	1	0	0
Gelsolin	GSN	FAF	35	1	1	12
Heterogeneous nuclear ribonucleoprotein D-like	HNRNPDL	Limb-girdle muscular dystrophy 1 G	15	1	1	4
Medin	MFGE8	Cerebrovascular dysfunction	115	1	1	29
Neurofilament heavy polypeptide	NEFH	CMT and ALS	1	1	0	1
Prion protein	PRNP	CJD, GSS, FFI	10	1	1	1
Receptor-interacting serine/threonine-protein kinase 1	RIPK1	Neuroinflammation	1	1	0	1
Solute carrier family 3 Member 2	SLC3A2	Lysinuric protein intolerance	57	1	0	8
Super oxide dismutase 1	SOD1	ALS	6	1	0	1
Transforming growth factor beta-induced	TGFBI	Corneal dystrophy	701	6	12	126
Tumor protein 53	TP53	Cancer	60	5	1	3
Transthyretin	TTR	Transthyretin amyloidosis	57	4	9	12
Tubulin alpha-1A chain	TUBA1A	Tubulinopathies	149	3	15	33

Column 1: Gene name. **Column 2:** Protein symbol. **Column 3:** Disease associated with protein. **Column 4:** Number of errors detected in transcripts that were derived from this gene. **Column 5:** Number of errors that generate mutant proteins identical to those seen in familial cases of protein aggregation diseases. **Column 6:** Number of errors that affect an amino acid (aa) known to be involved in disease, but mutate it to a different residue compared to the clinic. **Column 6:** Number of errors that increase the amyloid potential of these proteins as predicted by bioinformatic analysis (AmyPred-FRL).

Results

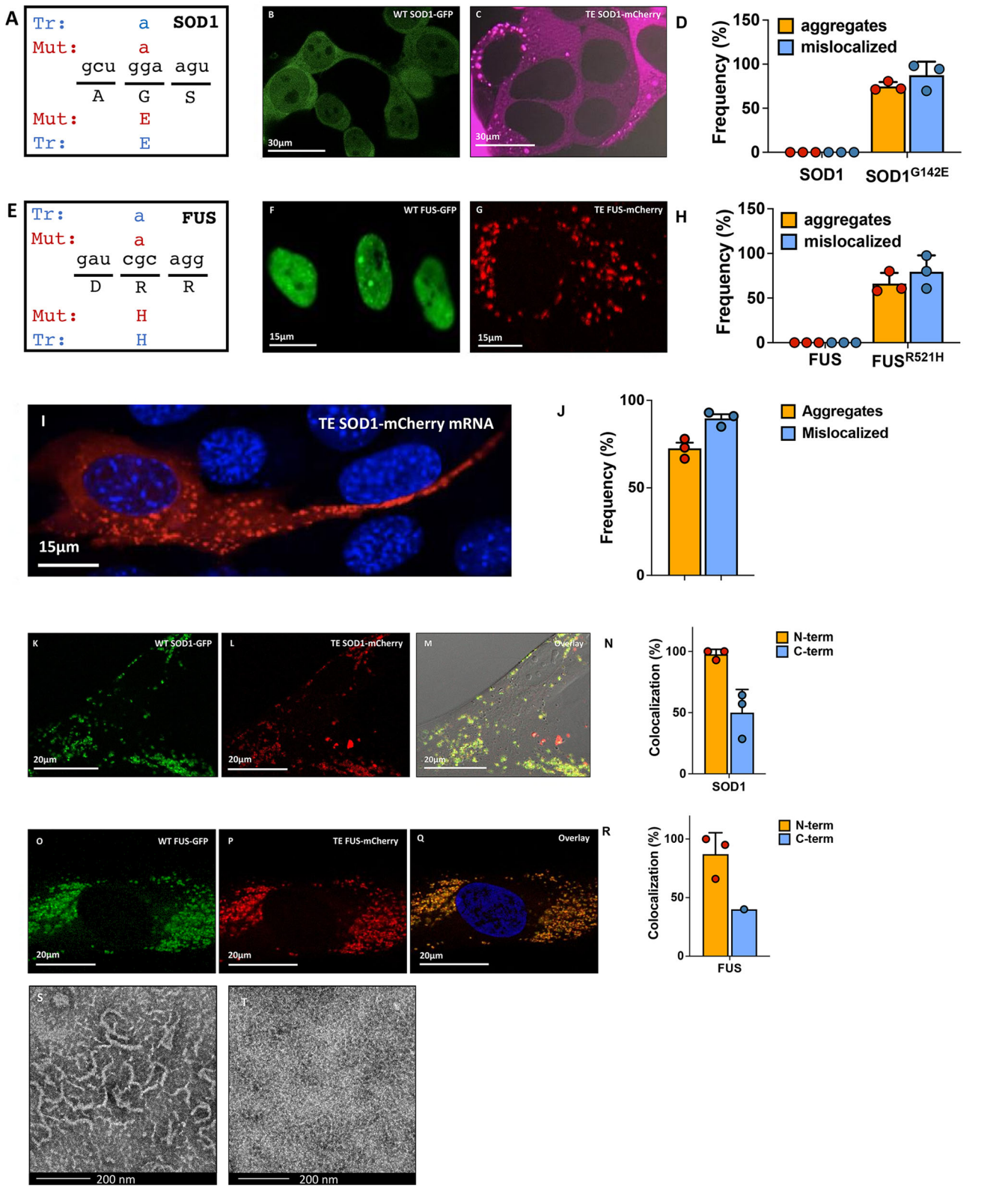
Transcript errors are ubiquitous in human cells

To test whether transcript errors give rise to amyloid or prion-like proteins, we probed the transcriptome of H1 human embryonic stem cells (H1 ESCs), brain organoids, neurons and fibroblasts with circ-seq, a massively parallel sequencing approach that uses consensus sequencing to enable high-fidelity RNA sequencing[17,19]. The brain organoids and neurons we sequenced were generated directly from the H1 ESCs (Supplementary Fig. 2), so that the genetic background between these models remained consistent and the results could be compared to each other. In addition, we sequenced the H1 ESCs at 300x coverage to generate a custom-made reference genome and ensure that single nucleotide polymorphisms or low-level mutations could be excluded from downstream analyses (Supplementary Fig. 3). In total, these sequencing efforts yielded >160,000 transcript errors that affected >11,000 genes across all three models (Fig. 1B). A complete list of the errors we detected can be found in the supplemental material attached to this publication. Each model displayed a similar error rate and spectrum, suggesting that the error rate of transcription is mostly independent of cellular fate, proliferation rate and differentiation status (Fig. 1C,D). We did observe a clear increase in A → G errors in the H1 ESCs cells though, which we previously found to reflect the impact of A to I RNA editing on the transcriptome[20].

Transcript errors create disease-associated proteins

We then used two approaches to determine if these errors result in amyloid or amyloid-like proteins: a literature-based approach and a bioinformatic approach. In our literature-based approach, we focused on 70 proteins that are directly implicated in various amyloid and amyloid-like diseases, including PRNP (CJD and Gerstman–Sträussler–Scheinker syndrome (GSS)[21]), APP (AD)[22]), SOD1, FUS (ALS)[23], and TTR (transthyretin amyloidosis) (for a list of selected proteins, see Table 1, for the full list, see Supplementary Table 1). Over the past 3 decades, hundreds of mutations have been identified in these proteins that cause familial cases of proteinopathies. In most cases, these mutations dramatically increase the amyloid and prion-like potential of the affected proteins. We reasoned that if transcript errors generate identical mutant proteins, they are likely to result in pathogenic proteins as well. To identify these errors, we cross-referenced the errors we detected with various databases that catalog germline mutations implicated in protein aggregation diseases, including Clinvar[24] and the Human Genome Mutation Database[25]. Of the 1936 errors that affected amyloid and amyloid-like proteins, we identified 38 errors that give rise to mutant proteins previously seen in the clinic. For example, 2 of the errors we detected generate mutant versions of the SOD1[26] and FUS protein[27], both of which were identified in familial cases of ALS (Table 1, Supplementary Table 1), while another error generated a mutant version of the human prion protein ($PRNP^{A133V}$) that causes GSS[28]. Other errors generated pathological versions of TTR (amyloidogenic transthyretin amyloidosis), CSTB (progressive myoclonic epilepsy), TGFBI (corneal dystrophy), APP (AD), CRYGD (coralliform cataracts), TP53 (cancer), Medin (natural aging), and others. In addition, we identified 75 errors that affect key amino acids directly implicated in disease, although these errors mutated these amino acids to a different residue compared to the clinic. For example, one of these errors generates a mutant version of PRNP ($PRNP^{V210A}$) that closely resembles a $PRNP^{V210I}$ mutation implicated in familial CJD[29] (both alanine and isoleucine are aliphatic amino acids). Similar errors were present in transcripts that encode APP, CSTB, HNRNPA1 (inclusion body myopathy with FTD), TGFBI, TP53, TTR and 10 other proteins (Table 1, Supplementary Table 1). A substantial portion of these errors are likely to affect the amyloid-like behavior of these proteins as well.

https://doi.org/10.1038/s41467-024-52886-2



Transcript errors create disease-associated amyloid-like proteins

To confirm that the errors we identified through our literature-based approach indeed result in proteins with amyloid or amyloid-like behavior, we selected two candidates for follow-up experiments. One of these errors generates a mutant version of SOD1 ($SOD1^{G142E}$, Fig. 2A–D) while the second error generates a mutant version of FUS (FUS^{R521H}, Fig. 2E–H). These mutant proteins were previously identified in familial cases of ALS[26,27]. We expressed these proteins in primary human fibroblasts, HEK293 cells and glioblastoma cells by lentiviral transfection (Fig. 2, Supplementary Fig. 4) and then imaged them by confocal microscopy. Consistent with the idea that transcription errors generate mutant proteins that display amyloid-like behavior, we found that both $SOD1^{G142E}$ and FUS^{R521H} aggregated in all 3 cell types, while the WT proteins did not. In addition, we found that the mutant SOD1 and FUS proteins were mislocalized. While WT FUS is

Fig. 2 | Transcription errors result in proteins with increased amyloid-like behavior. A A transcription error (Tr) was identified in the *SOD1* transcript that mimics a mutation (Mut) implicated in ALS. This error substitutes a guanine for an adenine base, resulting in a glycine (G) to glutamine (E) mutation at residue 142. **B** WT SOD1 is soluble and present throughout the cell, including the nucleus. **C** In contrast, $SOD1^{G142E}$ proteins form aggregates and are excluded from the nucleus. **D** Quantification of WT and mutant SOD1 aggregation and mislocalization. Depicted are the % of cells with aggregates or mislocalized proteins ($n = 3$ biological replicates). For aggregates, $P = 0.0015$, for mislocalization, $P = 0.0102$. **E** A transcription error was identified in the *FUS* transcript that mimics a mutation implicated in ALS. This error substituted a guanine for an adenine base, resulting in an arginine (R) to histidine (H) mutation at residue 521. **F** WT FUS is present in a soluble state in the nucleus, while FUS^{R521H} (**G**) forms aggregates outside of the nucleus. **H** Quantification of FUS aggregation and mislocalization ($n = 3$ biological replicates). Depicted are the % of cells with aggregates or mislocalized proteins. For aggregates, $P = 0.0108$, for mislocalization, $P = 0.0174$. **I**, **J** Transfection of cells with 500 ng of $SOD1^{G142E}$ mRNA resulted in cells with clearly visible mislocalized protein aggregates ($n = 3$ biological replicates). **K**–**M** When cells are transducted with lentiviral particles carrying both WT *SOD1* and $SOD1^{G142E}$ simultaneously, WT SOD1 is excluded from the nucleus and recruited into extranuclear aggregates. **N** Quantification of SOD1 colocalization with either N-terminal or C-terminal tags. **O**–**Q** WT and FUS^{R521H} co-expression in primary human fibroblasts, demonstrating that mutant and WT FUS co-localize in cytoplasmic aggregates ($n = 3$ biological replicates). **R** Quantification of FUS colocalization with either N-terminal or C-terminal tags. ($n = 3$ biological replicates). **S** WT SOD1 does not form fibers under TEM, but $SOD1^{G142E}$ does (**T**). TEM experiments were performed 3 times with similar results. $^{*}P < 0.05$. $^{**}P < 0.01$ according to a two-tailed unpaired t-test with Welch's correction. Data are presented as mean values ± SEM. Source data are provided as a Source Data file.

predominantly present in the nucleus (where it aids RNA splicing, gene expression and DNA repair[30]), the mutant protein was excluded from the nucleus and formed large punctate deposits throughout the cytoplasm (Fig. 2E–H). These observations complement similar results by others[30–34]. Similarly, SOD1 is normally distributed throughout the cytoplasm and the nucleus, but we found that $SOD1^{G142E}$ was excluded from the nucleus and formed large protein deposits in the cytoplasm (Fig. 2B–D). Importantly, nuclear exclusion and protein aggregation of FUS and SOD1 are key components of the pathology associated with ALS[35,36]. We observed the same mislocalization and aggregation when we mimicked transcriptional mutagenesis by transfecting 3T3 cells with mRNA from a $SOD1^{G142E}$ template. The protein aggregates generated by these mRNAs remained visible for at least 7 days (our latest time point in this experiment, Fig. 2I,J), indicating that they persist for an extended period after synthesis. Finally, we used lentiviral transduction to co-express WT and mutant SOD1 in the same cells and monitored their behavior. We found that when co-expressed with $SOD1^{G142E}$, WT SOD1 no longer distributed equally throughout the cells, but was excluded from the nucleus and assembled into the same amyloid-like deposits as $SOD1^{G142E}$ (Fig. 2J–N), suggesting that WT SOD1 was recruited by $SOD1^{G142E}$ and converted to an amyloid-like state. We made similar observations for WT and mutant FUS^{R521H} (Fig. 2O–R). WT FUS was almost always excluded from the nucleus in the presence of FUS^{R521H}, and sequestered in cytoplasmic deposits with FUS^{R521H}, although rare exceptions did occur (Supplementary Fig. 5). Consistent with the idea that $SOD1^{G142E}$ has amyloid-like properties, transmission electron microscopy (TEM) demonstrated that mutant SOD1 can form amyloid-like fibers in vitro (Fig. 2S,T). When taken together, these experiments provide an important proof of principle of the idea that transcription errors give rise to amyloid-like proteins. Moreover, because RNAPII constantly generates new mRNA molecules inside cells, and transcription by RNAPII is relatively error prone, (the error rate of transcription is approximately >100-fold higher than the mutation rate[37]), we conclude that transcription errors generate a continuous stream of amyloid and amyloid-like proteins in human cells.

Transcript errors create uncharacterized amyloid-like proteins

In addition to proteins directly connected to disease, we wondered whether transcription errors can also generate mutant proteins whose amyloid-like properties have not been characterized yet. To test this hypothesis, we used an unbiased bioinformatic approach to analyze the impact of errors on amyloid and amyloid-like proteins. First, we used AmyPred-FRL to analyze errors that affect amyloid- and prion-like proteins and found that 457 were predicted to increase their amyloid potential (Table 1, Supplementary Table 1). Second, we used PAPA[38] to analyze errors that affect proteins with prion-like domains. Although the *PRNP* gene encodes the canonical prion protein in humans, many proteins are now known to contain prion-like domains. Mutations in these domains can increase the prion-like behavior of these proteins in a wide variety of contexts, including proteotoxic diseases. For example, mutations in the prion-like domain of HNRNPA1 and HNRNPAB2 increase the pathogenic behavior of these proteins and can cause multisystem proteinopathies and ALS[39,40]. By applying this algorithm to our dataset, we found that 393 transcript errors are predicted to display increased amyloid- and prion-like behavior (Table 2, Supplementary Table 2).

Next, we extracted information from Prionscan[41], PLAAC[42] and the Amyloid Protein Database[43] to build a comprehensive database of proteins that have the potential to display amyloid and prion-like features. We then cross-referenced this database with the transcription errors we detected to identify errors that are likely to enhance these features. To test the accuracy of these predictions, we examined errors that affect the TP53 protein in greater detail. TP53 is an essential tumor suppressor protein involved in DNA repair, transcription, cellular senescence and apoptosis, and aggregates in 15% of human cancers[34]. With the bioinformatic tools described above we identified 5 transcription errors that are likely to increase the amyloid propensity of TP53: $TP53^{S149F}$, $TP53^{G245S}$, $TP53^{G279A}$, $TP53^{S315F}$ and $TP53^{P318L}$. When we mapped these mutations onto the crystal structure of TP53 we noticed that the S149F mutation (Fig. 3A) is located in a loop at the edge of the TP53 β-sandwich core (loop 146-WVDSTPPPGTR-156). Based on its location and the structural change it introduces (Fig. 3B,C), we predicted that this mutation may increase the amyloid propensity of the local peptide sequence (the 146-WVDSTPPPGTR-156 loop) and enhance the interaction between the β-sandwich cores of separate TP53 monomers, thereby leading to the assembly of the extended β-sheet structures that are characteristic of amyloid proteins. Mutations in the loop at the edge of the β-sandwich core of the TTR protein were previously shown to promote amyloid formation through a similar structure-based mechanism[44]. To test this hypothesis, we expressed the core domain (aa 92-292) of WT and mutant TP53 in bacterial cells and analyzed the behavior of these proteins by TEM. Consistent with our predictions, we found that $TP53^{S149F}$ indeed aggregated into large protein deposits, while WT TP53 did not (Fig. 3D,E). These aggregates displayed Congo-red birefringence under polarized light, a strong indicator of amyloid formation (Fig. 3F,G). We conclude that in addition to amyloid-like proteins directly implicated in disease, transcription errors can also give rise to uncharacterized mutant proteins with amyloid-like behavior.

Amyloid formation can be caused by a small number of errors

Next, we decided to test if $TP53^{S149F}$ can convert WT TP53 to an amyloid-like state, similar to $SOD1^{G142E}$ and FUS^{R521H}, and if so, how much $TP53^{S149F}$ was required to initiate this process. To answer this question, we added vanishing amounts of $TP53^{S149F}$ to a WT TP53 solution and found by TEM that 1% of $TP53^{S149F}$ (v/v) was sufficient to initiate the aggregation of the WT protein (Fig. 3H). We confirmed

Table 2 | Transcription errors affect proteins with prion-like domains

Gene	Protein	PrlD	Errors detected	Errors in PrlD	Prion potential ↑
Annexin A11	ANXA11	41-177	25	6	3
Cell Cycle Associated Protein 1	CAPRIN1	466-628	59	14	20
DEAD-Box Helicase 5	DDX5	531-614	64	3	14
DEAD-Box Helicase 17	DDX17	614-719	46	4	9
Ewing Sarcoma breakpoint region 1	EWSR1	1-256	68	21	17
Fused in Sarcoma	FUS	1-239	35	8	11
Heterogeneous nuclear ribonucleoprotein A0	HNRNPA0	206-305	23	7	3
Heterogeneous nuclear ribonucleoprotein A1	HNRNPA1	190-307	167	19	27
Heterogeneous nuclear ribonucleoproteins A2/B1	HNRNPA2B1	197-353	38	6	12
Heterogeneous nuclear ribonucleoprotein A3	HNRNPA3	207-378	36	6	14
Heterogeneous nuclear ribonucleoprotein D0	HNRNPD	195-260	53	3	9
Heterogeneous nuclear ribonucleoprotein H3	HNRNPH3	268-346	15	3	4
Heterogeneous nuclear ribonucleoprotein U	HNRNPU	683-825	132	6	15
Heterogeneous nuclear ribonucleoprotein U	HNRNPU	683-825	132	6	15
Heterogeneous nuclear ribonucleoprotein U-like 1	HNRNPUL1	531-766	105	9	17
Heterogeneous nuclear ribonucleoprotein U-like 2	HNRNPUL2	641-747	21	4	5
Interleukin enhancer-binding factor 3	ILF3	661-894	67	7	19
Nuclear receptor coactivator 1	NCOA1	908-1198	8	3	5
Nuclear factor of activated T-cells 5	NFAT5	999-1454	12	6	4
Nucleoporin 153	NUP153	1322-1453	23	4	4
Polyhomeotic Homolog 1	PHC1	7-104	103	8	5
R3H domain-containing protein 2	R3HDM2	394-714	8	3	3
SRY-Box Transcription Factor 2	SOX2	154-231	14	4	0
TAR DNA-binding protein 43	TARDBP	277-414	35	3	5
TRK-fused gene protein	TFG	223-400	24	8	6
Yip1 domain family member 5;	YIPF5	1-85	12	4	1
Zinc Finger MIZ-Type Containing 1	ZMIZ1	248-403	24	4	5

Column 1: Gene name. **Column 2:** Protein symbol. **Column 3:** Location of prion-like domain inside protein. **Column 4:** Number of errors detected in transcripts that were derived from this gene. **Column 5:** Number of errors that affect the prion-like domain. **Column 6:** The number of errors that increase the prion-like potential of these proteins as predicted by bioinformatic analysis (PAPA)

these findings in a dynamic light scattering experiment (Fig. 3I–K) that demonstrated that while WT TP53 was present at a size consistent with TP53 monomers, $TP53^{S149F}$ aggregated into deposits that were >100-fold larger in size. Moreover, when we added 2% (v/v) $TP53^{S149F}$ to the WT solution, we observed a disproportionate increase in TP53 aggregates that could only be explained by mutant-induced aggregation of WT proteins. Finally, we used atomic force microscopy (AFM) to characterize the cross-seeding behavior between WT and mutant TP53 further. First, we prepared a seeding solution of $TP53^{S149F}$ aggregates (Fig. 3L), which consists of both branched and helical structures, by sonication and centrifugation with an average particle size of 0.1 μm as determined by multi-angle light scattering (MALS) and AFM. Particles that are 0.1 μm in size are roughly equivalent to ~800-1000 molecules, a number that could be generated by the translation of 1 or a few mutant transcripts[45] (Fig. 3M). We then mixed these particles into the WT TP53 solution (Fig. 3N) at a 2% v/v ratio and observed a remarkable seed-dependent growth of WT TP53 fibers (Fig. 3O). When given sufficient incubation time, a 1:50 mixture of mutant:WT proteins created deposits that were visible to the naked eye (Fig. 3P). Given the size of these aggregates, these deposits must be constructed almost exclusively from WT proteins, with the mutant proteins serving as the initial seed.

To expand on these observations and ensure that this phenomenon is not caused by artifacts like AFM sample preparation (which involves drying samples on a mica surface), we developed a hanging drop method to characterize the seeding process in solution (Fig. 4A–C). First, we prepared a seeding solution of $TP53^{S149F}$ particles harvested from bacteria with an average size of 0.1 μm as determined by MALS and AFM (Supplementary Table 3). Then, we set up a 4 × 6 screening tray with a 1 ml reservoir solution that contains the protein buffer and an increasing concentration of NaCl (0.3, 0.5, 0.7, 0.8, 1.0, 1.2 M for columns 1–6, respectively). Finally, we added a 10 μl WT TP53 solution (60 μM) to a siliconized coverslip and placed a 1 μl drop of $TP53^{S149F}$ seed particles immediately adjacent to it at different concentrations (0, 1.2, 6, or 12 μM from row A to B, C and D). Over time, the protein drops on the coverslip shrink as a function of the NaCl concentration, gradually increasing the protein concentration. We reasoned that if the mutant seed particles display amyloid potential, this increasing concentration will eventually trigger the conversion of WT proteins to an amyloid state at the drop-drop interface and lead to localized fiber growth (Fig. 4C). Consistent with this idea, we observed robust growth of TP53 fibers under a light microscope at the WT:mutant interface, but not in the absence of the mutant protein (Fig. 4D, E). This rod-like material displayed strong birefringence under polarized light, which is suggestive of amyloid structures (Fig. 4F). Taken together, these biophysical experiments support the idea that transcription errors create amyloid proteins that can convert WT proteins to an amyloid state, which initiates the formation of large amyloid fibers and deposits. In addition, they suggest that a limited number of mutant transcripts is sufficient to initiate this process. For example, if 2% of $TP53^{S149F}$ proteins is sufficient to initiate the aggregation of WT TP53, then 2% of *TP53* transcripts encoding the $TP53^{S149F}$ mutation should be sufficient to initiate fiber formation as well. Similar thresholds were previously observed for other amyloid proteins, and it has been speculated that for prions, there may not be a safe dose at all[46].

https://doi.org/10.1038/s41467-024-52886-2

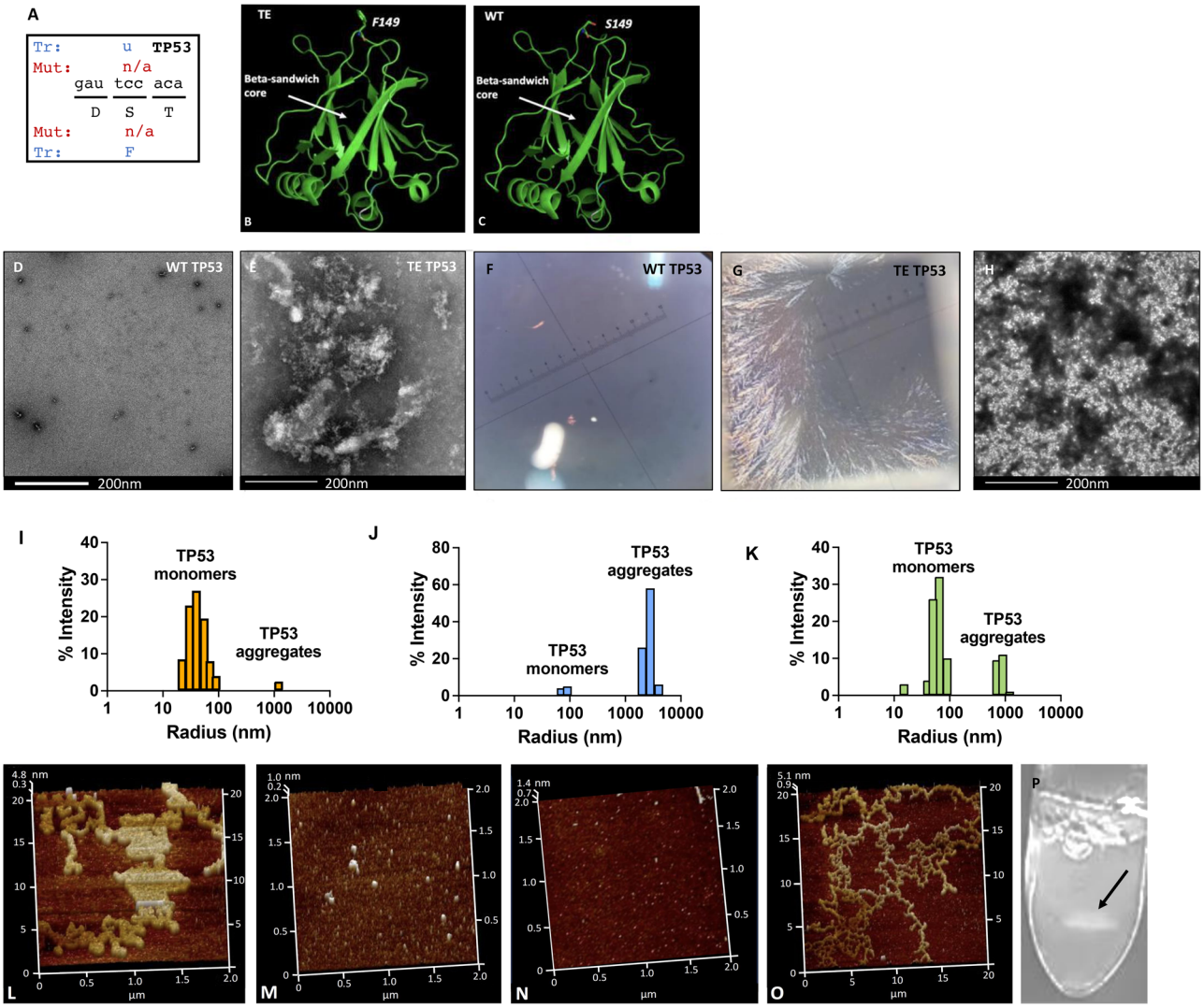


Fig. 3 | Biophysical examination of WT and mutant TP53. A A transcription error (Tr) was identified in a *TP53* transcript that substitutes a uracil for a cytosine base, resulting in a serine (S) to phenyl-alanine (F) mutation at residue 149. **B**,**C** Predicted structure of WT (**B**) and mutant TP53 (**C**). **D** Transmission electron microscopy showed little or no aggregates of WT TP53, while $TP53^{S149F}$ induces large protein aggregates (**E**). These experiments were performed 3–6 times with similar results. **F**,**G** Congo-red birefringence under polarized light indicates that $TP53^{S149F}$ forms amyloid fibrils (**G**), while WT TP53 does not (**F**). **H** After addition of 1% $TP53^{S149F}$ to a solution of WT TP53 (v/v), the WT solution generated countless aggregates. This experiment was performed 3 times with similar results. **I**–**K** Dynamic light scattering, which can be used to determine the radius of protein particles, indicates that WT TP53 is primarily in a monomeric form (**I**), while mutant TP53 consists of aggregates greater than 1000 nm (**J**). After 2% $TP53^{S149F}$ is added to a solution of WT TP53 (v/v), a large amount of TP53 aggregates emerges (**K**). **L** $TP53^{S149F}$ aggregates into a variety of structures. **M** TP53 aggregates were sonicated to create a seed solution of particles that are around 0.1 µm in size, which equates to 800-1000 proteins. (**N**) WT TP53 solution shows no apparent aggregation; (**O**) Adding the $TP53^{S149F}$ amyloid seed solution to WT TP53 in a 1:100 ratio induced fibril growth. **P** Protein aggregates created by mutant TP53 form spontaneously and can be seen by the naked eye (arrow).

Human cells possess various mechanisms to counteract protein aggregation though, including molecular chaperones, autophagy and the ubiquitin-proteasome system. It is possible then, that the threshold for protein aggregation differs between in vitro and in vivo scenarios. To determine the threshold for intact cells, we generated mRNA molecules from WT and mutant *SOD1* templates, mixed them together in various ratios and transfected them into primary human fibroblasts and 3T3 cells. First, we transfected cells with a 50:50 ratio of WT and mutant transcripts (250 ng:250 ng WT:mutant mRNA) and then gradually lowered the mutant ratio to 10% (450 ng:50 ng WT:mutant mRNA). We found that co-aggregation still occurred at a ratio of 10% mutant RNA after 7 days in culture, indicating that in human cells, the threshold for aggregation is at least 10% (Fig. 4G–K). Mixtures with only 2% mutant proteins did not produce sufficient mCherry signal to confidently visualize co-aggregation.

DNA damage creates transcripts with identical errors

With this threshold in mind, we decided to test if it is possible for 10% of transcripts to carry the same transcription error. Importantly, it was previously shown that DNA damage can provoke the same mistake by RNA polymerase II (RNAPII) during multiple rounds of transcription[14] (Fig. 5A), so that up to 50% of transcripts can carry the same transcription error[47–49]. These studies were primarily performed on DNA repair deficient cells though, using a single DNA lesion placed on a plasmid. As a result, it is unclear how well these findings translate to a WT genome carefully wrapped in chromatin that is actively surveyed by DNA repair. Therefore, we designed a single cell sequencing approach to examine the impact of DNA damage on transcriptional mutagenesis.

First, we treated quiescent mouse neural stem cells (NSCs) that were derived from the hippocampus for 1 h with MNNG, a powerful mutagen that randomly generates O^6-methyl-guanine adducts (O^6-me-

https://doi.org/10.1038/s41467-024-52886-2

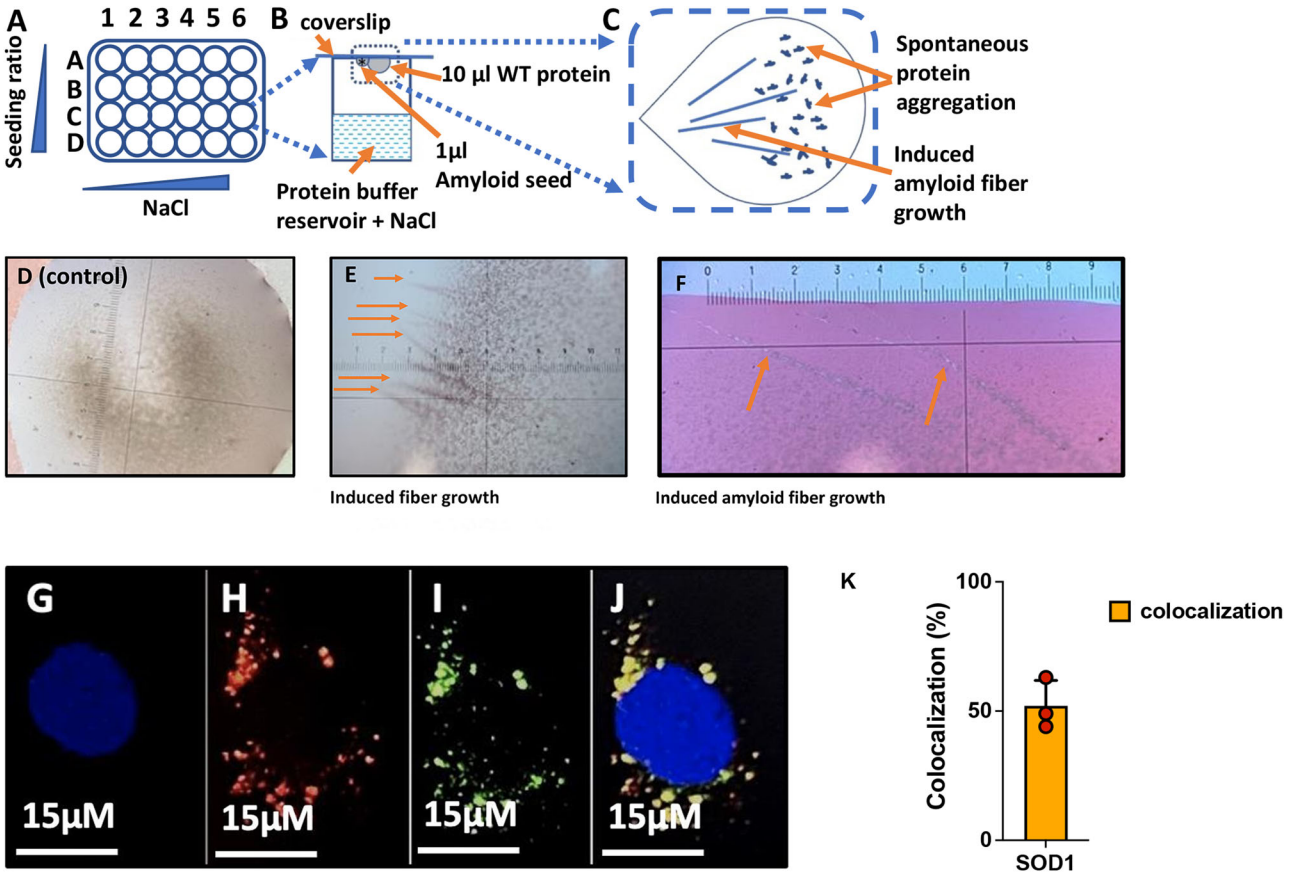


Fig. 4 | A hanging drop method and mRNA transfections to assess WT:mutant ratios required for protein aggregation. A 4 × 6 screening tray was set up with a 1 ml reservoir that contains protein buffer and an increasing concentration of NaCl. **B** A 10 µl WT TP53 solution (60 µM) was then added to a siliconized coverslip and a 1 µl drop of $TP53^{S149F}$ seed particles was placed immediately adjacent at decreasing concentrations. **C** If the mutant seed particles have amyloid potential, this event will trigger conversion of WT proteins at the drop-drop interface and lead to localized fiber growth **D** If no $TP53^{S149F}$ is provided as seeding material, no fiber-like material forms in the WT TP53 solution. **E** However, if $TP53^{S149F}$ seeding material is provided, fiber-like material grows out of the WT solution. **F** These fibers show strong birefringence under polarized light, suggestive of amyloid structures. **G**–**J**. Primary fibroblasts were transfected with 90% WT and 10% mutant $SOD1^{G142E}$ transcripts display co-localized WT and mutant proteins inside protein aggregates **G** DAPI. **H** $SOD1^{G142E}$-mCherry. **I** WT SOD1-eGFP. **J** Overlay of **G**–**I**. These experiments were performed 3 times with similar results. **K** Quantitation of **G**–**J** ($n = 3$ biological replicates), data is presented as mean values ± SEM. Source data are provided as a Source Data file.

G)[50]. We chose hippocampal stem cells for these experiments because they are directly implicated in amyloid diseases[51], and O^6-me-G adducts because they play an important role in human brain cancers[52,53] and were recently implicated in the pathology of female patients with AD[54]. We performed these experiments on non-dividing NSCs (Supplementary Fig. 6), so that the O^6-me-G lesions we induced would not be fixed into mutations during DNA replication (a common experimental setup to prevent mutations from confounding transcription error measurements[14,47–49,55]). After MNNG treatment, we provided the cells with fresh medium, and let them recover for increasing periods of time. We then sequenced the transcriptome of single cells at different timepoints (Fig. 5B) to identify transcription errors that occurred in at least 10% of transcripts from a gene, with a minimum of 40 unique transcripts sequenced. These parameters also prevent direct damage to RNA molecules from affecting our measurements, because it is unlikely that this damage will affect the same nucleotide on multiple RNA molecules.

We found that MNNG treatment resulted in a > 40-fold increase in transcripts with identical errors after 16 h of recovery time (Fig. 5C). The vast majority of these events (which we labeled pseudo-alleles for their ability to generate WT and mutant transcripts) were C → U errors, the most common error induced by O^6-me-G lesions. Notably, no increase was detected in G → A errors, which would have occurred if O^6-me-G lesions had been fixed into CG:TA mutations, demonstrating that our experiment was not confounded by conventional mutagenesis. Consistent with this idea, we found that G → A errors did arise in dividing cells (Supplementary Fig. 6). In most cases, pseudo-alleles gave rise to 10–20% of mutant transcripts (Fig. 5D), either meeting or exceeding the threshold required for protein aggregation in vitro and in vivo.

Cells exposed to DNA damage experience proteotoxic stress

Consistent with the idea that the errors generated by these pseudo-alleles cause protein misfolding and aggregation, we found that treated cells displayed a substantial increase in markers for misfolded proteins and proteotoxic stress, particularly at the time point that the errors reached their peak (Fig. 5E). Accordingly, human cells that display error prone transcription[56] display increased protein aggregation as well (Supplementary Fig. 7). We further note that the number of pseudo-alleles rose over time as more and more genes were transcribed, and were still present 16 h after exposure, indicating that transcriptional mutagenesis is not only abundant after exposure, but also long-lasting, even in cells capable of DNA repair.

Interestingly, loss of DNA repair is increasingly implicated in amyloid diseases[57]. For example, it was recently reported that the promoter of the gene that encodes the main DNA repair protein for O^6-me-G lesions in human cells (MGMT[53]) is hypermethylated in female patients with AD[54], suggesting that in these patients, pseudo-alleles

https://doi.org/10.1038/s41467-024-52886-2

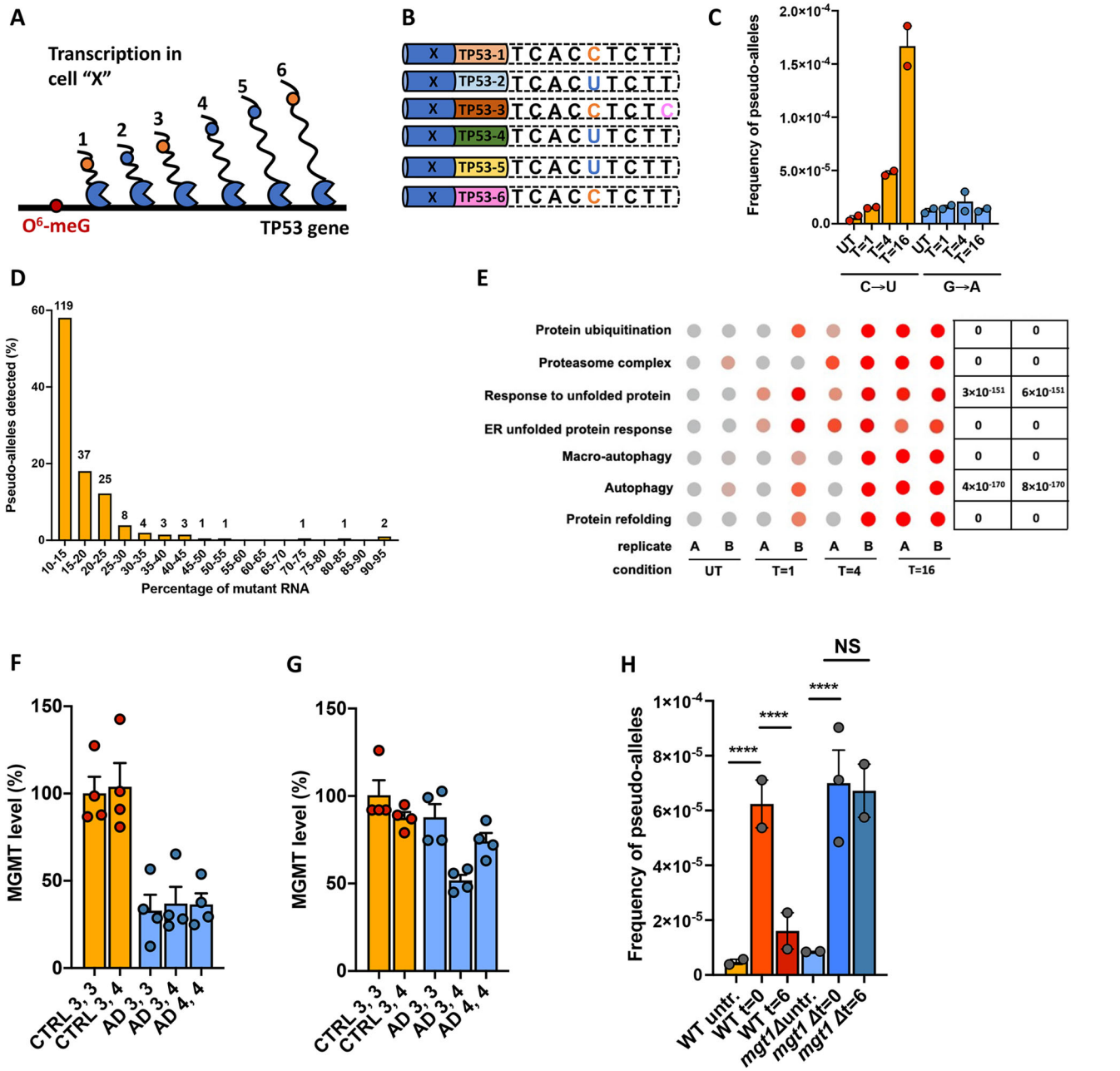


Fig. 5 | DNA damage and off-target RNA editing affect the fidelity of transcription. A If DNA damage results in multiple rounds of error prone transcription, then multiple transcripts in a single cell should carry identical errors. **B** When these transcripts are captured and tagged with UMIs, they can be grouped together, and their sequences can be compared to each other to search for identical errors that occur in multiple transcripts. In contrast, sequencing errors or RNA damage will only be present in one transcript. Blue bar: cell-specific barcode. Multi-colored bar: transcript UMI. Blue base: WT. Orange base: transcription error. Pink base: Sequencing error/RNA damage **C** C → U pseudo-alleles emerge after MNNG treatment of mouse neural stem cells, while G → A errors (which would indicate conventional mutagenesis is occurring as well) do not ($n = 2$ biological replicates). *$P < 0.05$, ****$P < 0.0001$ according to a Chi-squared test with Yates' continuity correction. **D** Ratio of WT:mutant mRNAs identified. Only alleles with more than 10% mutant mRNAs are depicted. **E** Dot plots of single cell gene expression profiles grouped by GO-terms indicate markers of proteotoxic stress are elevated in treated cells, particularly at 16 h, when the transcript error rate is the highest. Significance was ascertained by ANOVA test. FDR= False Discovery Rate. **F** MGMT levels are decreased in all females with AD compared to control females with an *APOE3,3* genotype. For AD 3,3, $P = 0.0022$, for AD 3,4, $P = 0.0034$, for AD 4,4, $P = 0.0022$. $n = 4$ biological replicates, a two-tailed unpaired t-test with Welch's correction. **G** MGMT levels are not decreased in males with AD, except for those with a *APOE3/APOE4* genotype, $P = 0.0066$. $n = 4$ biological replicates, a two-tailed unpaired t-test with Welch's correction. **H** Loss of *MGT1*, the yeast homolog of *MGMT* allows O^6-me-G lesions to remain on the genome, resulting in greatly increased numbers of pseudo-alleles over time. $n = 2$ biological replicates. ****$P < 0.0001$ according to a Chi-squared test with Yates' continuity correction. Source data are provided as a Source Data file.

could be present for an extended period of time. To test this hypothesis, we first confirmed that female AD patients indeed display reduced MGMT expression by Western blots. Consistent with a previous study, males did not display this trend (Fig. 5F,G, Supplementary Fig. 8, Supplementary Table 4). To mimic the impact of reduced MGMT expression on human cells, we then deleted the yeast homolog of *MGMT* (*mgt1*) in the budding yeast *S. cerevisiae*, arrested them in G1 with α-mating factor (Supplementary Fig. 9) and exposed them to MNNG. Similar to human cells, we found that WT yeast cells displayed an increase in pseudo-alleles immediately after exposure, which

https://doi.org/10.1038/s41467-024-52886-2

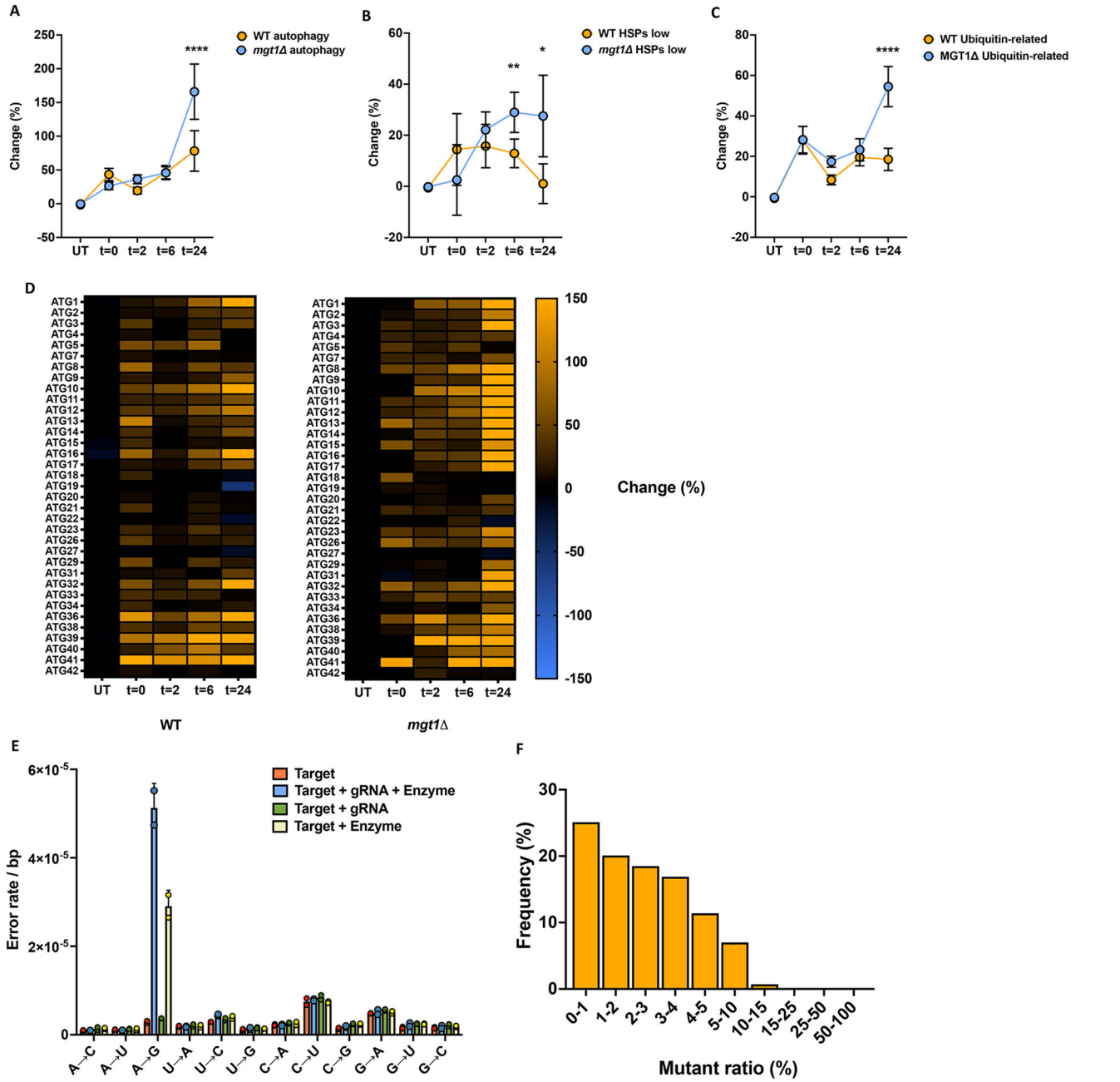


Fig. 6 | DNA damage induces markers of proteotoxic stress. Consistent with the idea that these errors result in misfolded proteins, these cells displayed markers of proteotoxic stress, including upregulated autophagy genes (**A**), heat shock proteins (**B**) and proteins implicated in the ubiquitin-proteasome system (**C**). Depicted in figure **A** and **C** is the average percentage change for all autophagy and ubiquitination-related genes identified by bulk RNA-seq. The genes depicted in **B** have been separated from several heat shock proteins that displayed unusually large increases in transcript levels (Supplementary Fig. 9, Supplementary Table 4). $n = 3$ biological replicates, for autophagy $P < 0.0001$, for heat-shock proteins $P = 0.0016$ at 6 h and $P = 0.0287$ at 24 h. For ubiquitin $P < 0.0001$. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{****}P < 0.0001$ according to a two-tailed paired t-test with Welch's correction. **D**. Heat map of autophagy genes detected in WT and mutant cells. **E** Error spectrum of human cells after transformation with plasmid that carries an editing target, the gRNA required to edit the target, and the editing enzyme. If the editing enzyme is present, large numbers of A to I editing events (A to G errors) were observed. $n = 2$ biological replicates. **F** Percentage of editing events that generate mRNAs with various mutant:WT ratios. Data are presented as mean values ± SEM. Source data are provided as a Source Data file.

declined after DNA repair was able to remove these lesions from the genome (Fig. 5H). However, in the absence of Mgt1p, the pseudoalleles remained on the genome, causing transcriptional mutagenesis for an extended period of time.

These observations confirm our recent findings, which show that Mgt1p removes 90% of DNA lesions within a 6-h timespan, and that no mutations arise in non-dividing cells treated with MNNG[58]. Similar to neural stem cells, treated yeast cells displayed increased expression of autophagy genes, molecular chaperones and components of the ubiquitin-proteasome system, indicating that they are under proteotoxic stress (Fig. 6A–D, Supplementary Fig. 10, Supplementary Table 5–7). Consistent with the idea that these markers are upregulated due to transcript errors, we previously found that yeast cells that display error prone transcription also show increased markers of proteotoxic stress[13]. In contrast, markers associated with translation (which is inhibited in times of proteotoxic stress) were downregulated (Supplementary Fig. 11, Supplementary Data 1). *mgt1Δ* cells showed a prolonged response of these markers, consistent with the prolonged

https://doi.org/10.1038/s41467-024-52886-2

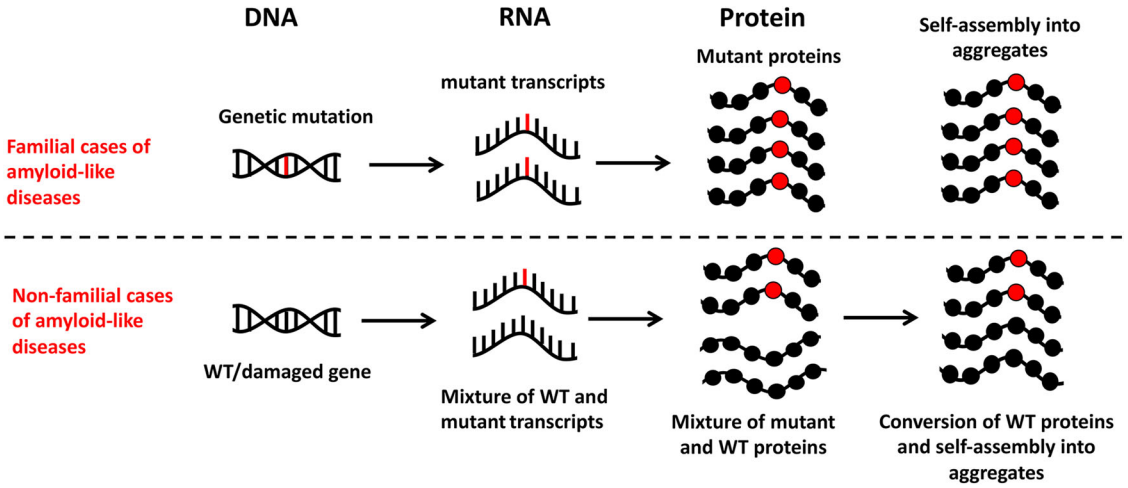


Fig. 7 | Model for the contribution of transcription errors to non-familial cases of disease. Familial cases of protein aggregation diseases are caused by genetic mutations that generate mutant proteins with increased amyloid or amyloid-like behavior. In non-genetic cases, identical mutant proteins (and potentially unique mutant proteins) are generated by non-genetic mutations such as transcription errors. Over time, these proteins convert WT proteins to an amyloid state, leading to a later onset of amyloid and amyloid-like diseases compared to familial cases. Because these amyloid and amyloid-like proteins are generated by mutations that are only present in transcripts though, they have thus far gone undetected in the clinic.

presence of pseudo-alleles on their genome (Supplementary Figs. 5I–K, 11, Supplementary Data 1).

RNA editing creates mutant transcripts with identical errors

In addition to transcription errors, it has been proposed that other molecular mistakes could result in amyloid and amyloid-like proteins as well, including off-target RNA editing[15]. One of the best-known examples of RNA editing in the animal kingdom is seen in cephalopods, where ADARs edit adenine to inosine (A to I) in a sequence-specific manner[59], an event that can be monitored by circ-seq as apparent A → G errors[20]. Editing in cephalopods is particularly abundant in neuronal tissues, while non-neuronal tissues display limited editing[59,60]. Consistent with this idea, we detected abundant A → G errors in the optic lobe and stellate ganglia, and relatively little in the gills, which we were able to pick up at both a high and low frequency (<1%), indicating that circ-seq can detect naturally occurring editing events at levels similar to off-target editing (Supplementary Fig. 12). Importantly, RNA editing tools are increasingly thought of as a tool to treat symptoms of disease[61]. To test whether RNA editing tools designed in the lab can result in off-target editing and the production of mutant proteins, we expressed the deaminase domain from human ADAR2, linked to gRNA through an optimized bacteriophage λN protein- BoxB hairpin[62,63] that was specifically designed to edit the *ATP1a3* transcript in human cells, and monitored off-target editing. We found that these editors display a substantial amount of off-target editing, whether a guide RNA is present or not (Fig. 6E). These events resulted in large numbers of rare (<2%) and common (>2%) mutant RNAs (Fig. 6F), suggesting that these editors have the potential to produce large amounts of mutant proteins in human cells, potentially increasing the risk for protein aggregation.

Discussion

To identify the molecular mechanisms that underpin human aging and understand how these mechanisms drive age-related pathology, it will be essential to determine how amyloid and amyloid-like proteins are generated. Here, we demonstrate that transcript errors represent one of these mechanisms. Although transcript errors are transient events, amyloid and amyloid-like proteins are characterized by their ability to replicate themselves by converting WT proteins to an amyloid-like state. Thus, even a transient event like a transcription error could trigger large-scale protein aggregation as a result of this self-templating mechanism.

One of the most intriguing observations from our experiments is that transcript errors generate mutant proteins that are already known to cause familial cases of amyloid disease. This observation suggests a unified mechanism for the development of familial and non-familial cases of human proteinopathies, as both could be caused by identical mutant proteins, only the mechanism by which the proteins are generated is different (Fig. 7). Consistent with this hypothesis, it was previously shown that aggregates of tau have identical structures in both familial and non-familial cases of AD[64], suggesting that they were initiated by identical mutant proteins.

However, we also detected a transcription error that generated an amyloid variant of the TP53 protein that had not been observed before, suggesting that transcription errors can also generate amyloid proteins that we are currently unaware of. Because these mutant proteins are likely to affect cellular proteostasis as well, it will be crucial to determine how many of these proteins exist, and how frequently they arise.

For these mutant proteins to cause long-term aggregation, transcription errors need to create enough mutant proteins to overcome the cellular protein quality control machinery and convert WT proteins to an amyloid state. Our in vivo experiments suggest that this threshold is breached when 10% of mRNA molecules encode an identical mutant protein, while our in vitro experiments suggest that theoretically, this limit could be as low as 1%. For example, a reduced threshold could occur in aging cells, which are known to display reduced protein quality control[65]. It is also likely that the threshold differs between proteins, or between mutations in the same protein. For example, some mutations could cause more aggressive conversion of WT proteins than others, thereby altering the threshold required for long-term protein aggregation.

Regardless, we show here that a 10% threshold is routinely breached as a result of DNA damage, a ubiquitous feature of aging cells that is closely associated with protein misfolding diseases. For example, farmers that are exposed to the DNA damaging pesticides rotenone and paraquat have an increased risk for developing AD and PD[66,67], while the DNA damaging agent methylazoxymethanol (MAM) is suspected of being the pathogenic agent responsible for Guam ALS/Parkinsonism-Dementia Complex, a disease that is characterized by protein aggregation and a variety of neurological symptoms[68–70]. Importantly, we recently discovered that MAM induces enough transcription errors to breach the >10% threshold in mouse neural stem

cells[71]. Although our experiments focused solely on pseudo-alleles created by O^6-me-G, it should be noted that other forms of DNA damage can generate pseudo-alleles as well[47–49], including oxidative DNA damage[14]. Thus, it is likely that numerous forms of DNA damage can trigger protein aggregation through a similar mechanism.

It may even be possible to use the data generated here to estimate the number of cells that carry an amyloid pseudo-allele. A survey of Clinvar and HGMD indicates that mutagenesis of at least 1038 bases per genome (covering 25 genes and 20 common amyloid diseases) could result in a pathogenic amyloid protein. Meanwhile, our single cell sequencing data shows that pseudo-alleles causing >10% mutant mRNA are present at a frequency of $\sim 5 \times 10^{-6}$/bp in human cells, or once every 200,000 bases. Thus, 1 out of every 200 cells could carry an amyloid pseudo-allele at any given time. Because pseudo-alleles are constantly lost and created (due to ongoing DNA repair and DNA damage) and neurons are unusually long-lived cells, we suspect that almost every neuron will carry a pseudo-allele at some point during their lifetime. The number of cells that carry a pseudo-allele could also change due to a variety of factors, including changes in DNA repair capacity, exposure to environmental pollutants, or medical interventions like chemotherapeutic treatments or engineered RNA editors. For example, we found that after one treatment of MNNG, the number of pseudo-alleles increased almost 10-fold, indicating that as many as 1:20 cells could carry an amyloid pseudo-allele. These observations demonstrate the potential of DNA damage to generate large amounts of identical mutant proteins without the need to induce mutations, as the damage itself is sufficient. It will be important to test whether real-life exposures result in similar increases in transcriptional mutagenesis. In addition, it would be exciting to detect these proteins directly, potentially by mass-spectrometry. Because DNA damage is randomly distributed across the genome though, each cell is likely to carry a different set of pseudo-alleles, making the mutant proteins present in one cell exceedingly rare compared to the WT proteins present in surrounding cells. Current mass-spectrometry technology is not yet capable of detecting these rare events, but single-cell proteomics, or increased sensitivity of targeted mass-spectrometry approaches may make these experiments possible in the future.

Consistent with a role for DNA damage in protein misfolding diseases, it is now increasingly recognized that loss of DNA repair can exacerbate amyloid diseases as well. For example, it was recently shown that the DNA repair gene *MGMT* is hypermethylated in female AD patients[54]. When we mimicked this phenomenon in yeast, we found that loss of the yeast homolog of *MGMT* allows pseudo-alleles to persist on the genome for extended periods of time, creating vast amounts of mutant proteins through transcriptional mutagenesis and a prolonged presence of markers associated with reduced proteostasis. It has long been known that females are at a greater risk for AD compared to males, and our data suggests that reduced MGMT expression, followed by extended transcriptional mutagenesis, could help explain the sexual dimorphism of AD[72].

Besides DNA damage, the fidelity of transcription can also be altered by other variables. For example, we previously found that the error rate of transcription increases with age in yeast[13] and flies, and is also affected by epigenetic markers, cell type and genetic context[12,20]. Although these variables may not increase the ratio of mutant:WT RNA to a potentially pathogenic level, it is important to note that transcription errors may not need to create highly specific amyloid proteins to promote amyloid diseases. Surprisingly, we previously found that random transcription errors can affect protein aggregation as well. Because the primary impact of mistakes in protein coding sequences is protein misfolding[73], random errors tend to create a cache of misfolded proteins that affect the entire proteome. Although the vast majority of these misfolded proteins are relatively benign, and do not give rise to an amyloid version of a critical protein, they do need to be degraded by the same protein quality control machinery as pathogenic proteins. As a result, random errors (like those created by an error prone RNAPII, Supplementary Fig. 7) can create enough misfolded proteins to overwhelm the protein quality control machinery, which then allows pathological amyloid proteins to evade degradation and seed aggregates[13]. Thus, transcription errors may not only generate highly specific amyloid and prion-like proteins, as we demonstrate here, they may also generate the conditions that allow these proteins to evade the protein quality control machinery and initiate aggregation. In this context, it is interesting to note that the speed of transcription increases with age in mammals and multiple model organisms[74]. Importantly, the speed of DNA and RNA polymerases is directly related to their fidelity, with higher speeds resulting in lower fidelity, and lower speeds resulting in higher fidelity, an evolutionary trade-off that is essential for the fitness of species[75]. When combined with the data presented here, these observations suggest that the increased speed of RNA polymerases in aging cells could lead to increased error rates, resulting in enhanced transcriptional mutagenesis and protein aggregation in aging organisms. Indeed, we previously found that highly expressed genes have higher error rates compared to rarely expressed genes in human cells[20] and that aged flies[56] and yeast cells[13] display increased error rates of transcription. Because lowering the speed of RNAPII was shown to increase their lifespan, it would be interesting to test whether this intervention increases the fidelity of transcription.

When taken together, these observations make a compelling case for the idea that transcription errors could play a role in the progression of multiple diseases caused by protein aggregation. If so, we expect this role to depend on a wide variety of variables, including the pathophysiology of the disease itself, the aggressiveness with which mutant proteins convert WT proteins to an amyloid state, and the error rate of RNAPII on critical bases. Because there are many variables, it is likely that the contribution of transcription errors to amyloid diseases varies from disease to disease. For example, in patients with non-familial cases of AD, errors have been detected in transcripts that code for the APP and UBB protein[76,77], creating peptides that are part of the amyloid plaques that characterize the disease. Importantly, these errors occur at GAGA repeats that are highly error-prone[12] and seem to be made especially frequently in hippocampal neurons[20], which are the primary target of the disease. Thus, multiple variables suggest that transcription errors could play an important role in non-familial cases of AD. In contrast, GSS is almost exclusively an inherited disorder, suggesting that transcription errors may have a limited impact on disease progression. These considerations underline that it will be important to use human cell lines, animals and patient samples to conclusively test the impact of transcriptional mutagenesis on the progression of amyloid diseases. Human cell lines that display error prone transcription (Supplementary Fig. 7) could play an important role in this process, as do human patients that carry error prone RNA polymerases[56]. Most importantly though, it will be critical to develop technology and animal models capable of dissecting the impact of transcriptional mutagenesis on disease progression while ruling out confounding factors. If successful, these developments could provide a unified mechanism for the etiology of various diseases that are currently endemic in our society, by demonstrating that both genetic and non-genetic cases of protein misfolding diseases are caused by mutant proteins, only the mechanism by which they are created is different (Fig. 7).

Methods

Ethics

All applicable international, national, and/or institutional guidelines for the care and use of previously published animal or human samples were followed. Stem cell work was approved by USC under SCRO Protocol #2019-3. All human subjects provided informed consent and human tissue use was approved through IRB protocol #UP-20-00014-EXEMPT.

H1 hESC cell culture

H1 hESCs were purchased from WiCell in Wisconsin (WA01) and cultured in TeSR medium (Stem Cell Technologies, 100-0276, 100-1130) in Matrigel-coated 10 cm plates (Corning, #354277). Cells were grown at 5% O_2 tension to better mimic the conditions inside the human body and reduce oxidative damage as a result of non-normoxic conditions. To passage cells and prior to collection of RNA and DNA, cells were gently treated with 2 μg/mL Dispase (Stemcell Technologies, #07913) mixed with DMEM/F12 (Thermo Fisher Scientific, #11320033), washed with PBS and scraped off the plate using a glass pipette. DNA and RNA were then isolated with standard phenol chloroform and Trizol methods.

Library construction and sequencing

Library preparation 1100 ng of enriched mRNA was fragmented with the NEBNext RNase III RNA Fragmentation Module (E6146S) for 25 min at 37 °C. RNA fragments were then purified with an Oligo Clean & Concentrator kit (D4061) by Zymo Research according to the manufacturer's recommendations, except that the columns were washed twice instead of once. The fragmented RNA was then circularized with RNA ligase 1 in 20 μl reactions (NEB, M0204S) for 2 h at 25 °C after which the circularized RNA was purified with the Oligo Clean & Concentrator kit (D4061) by Zymo Research. The circular RNA templates were then reverse transcribed in a rolling-circle reaction by first incubating the RNA for 10 min at 25 °C to allow the random hexamers used for priming to bind to the templates. Then, the reaction was shifted to 42 °C for 20 min to allow for primer extension and cDNA synthesis. Second strand synthesis and the remaining steps for library preparation were then performed with the NEBNext Ultra RNA Library Prep Kit for Illumina (E7530L) and the NEBNext Multiplex Oligos for Illumina (E7335S, E7500S) according to the manufacturer's protocols. Briefly, cDNA templates were purified with the Oligo Clean & Concentrator kit (D4061) by Zymo Research and incubated with the second strand synthesis kit from NEB (E6111S). Double-stranded DNA was then entered into the end-repair module of RNA Library Prep Kit for Illumina from NEB, and size selected for 500–700 bp inserts using AMPure XP beads. These molecules were then amplified with Q5 PCR enzyme using 11 cycles of PCR, using a two-step protocol with 65 °C primer annealing and extension and 95 °C melting steps. Sequencing data was converted to industry standard Fastq files using BCL2FASTQv1.8.4.

Error identification

We have developed a robust bioinformatics pipeline to analyze circ-seq datasets and identify transcription errors with high sensitivity (see code availability)[12,17]. First, tandem repeats are identified within each read (minimum repeat size: 30nt, minimum identity between repeats: 90%), and a consensus sequence of the repeat unit is built. Next, the position that corresponds to the 5′ end of the RNA template is identified (the RT reaction is randomly primed, so cDNA copies can start anywhere on the template) by searching for the longest continuous mapping region. The consensus sequence is then reorganized to start from the 5′ end of the original RNA fragment, mapped against the genome with tophat (version 2.1.0 with bowtie 2.1.0) and all non-perfect hits go through a refining algorithm to search for the location of the 5′ end before being mapped again. Finally, every mapped nucleotide is inspected and must pass 5 checks to be retained: (1) it must be part of at least 3 repeats generated from the original RNA template; (2) all repeats must make the same base call; (3) the sum of all qualities scores of this base must be >100; (4) it must be >2 nucleotides away from both ends of the consensus sequence; (5) each base must be covered by >100 reads with <1% of these reads supporting a base call different from the reference genome. This final step filters out polymorphic sites and intentional potential RNA-editing events. For example, if a base call is different from the reference genome, but is present in 50 out of 100 reads, it is not labeled as an error but as a heterozygous mutation. A similar rationale applies to low-level mutations and RNA editing events. These thresholds were altered to detect different types of editing events, including common editing events. Each read containing 1 or more mismatches is filtered through a second refining and mapping algorithm to ensure that errors in calling the position of the 5′ end cannot contribute to false positives. The error rate is then calculated as the number of mismatches divided by the total number of bases that passed all quality thresholds.

Brain organoid culture and generation

H1 ESC colonies were maintained with daily media change in mTeSR (STEMCELL Technologies, #85850), supplemented with a final concentration of 5 ՝M XAV-939 (STEMCELL Technologies, #72672) on 1:100 geltrex (GIBCO, #A1413301) coated tissue culture plates (CELLTREAT, #229106) and passaged using ReLeSR (STEMCELL Technologies, #100-0484). Cells were maintained below passage 50 and periodically karyotyped via the G-banding Karyotype Service at Children's Hospital Los Angeles. To generate dorsally patterned forebrain organoids, we modified the method previously described in Kadoshima et al.[78]. Briegly, on day 0, feeder-free cultured human PSCs, 80–90% confluent, were dissociated to single cells with Accutase (Gibco), and 9000 cells per well were reaggregated in ultra-low cell-adhesion 96-well plates with V-bottomed conical wells (sBio PrimeSurface plate; Sumitomo Bakelite) in Cortical Differentiation Medium (CDM) I, containing Glasgow-MEM (Gibco), 20% Knockout Serum Replacement (Gibco), 0.1 mM Minimum Essential Medium non-essential amino acids (MEM-NEAA) (Gibco), 1 mM pyruvate (Gibco), 0.1 mM 2-mercaptoethanol (Gibco), 100 U/mL penicillin, and 100 μg/mL streptomycin (Corning). From day 0 to day 6, ROCK inhibitor Y-27632 (Millipore) was added to the medium at a final concentration of 20 μM. From day 0 to day 18, Wnt inhibitor IWR1 (Calbiochem) and TGFβ inhibitor SB431542 (Stem Cell Technologies) were added at a concentration of 3 μM and 5 μM, respectively. From day 18, the floating aggregates were cultured in ultra-low attachment culture dishes (Corning) under orbital agitation (70 rpm) in CDM II, containing DMEM/F12 medium (Gibco), 2 mM Glutamax (Gibco), 1% N2 (Gibco), 1% Chemically Defined Lipid Concentrate (Gibco), 0.25 μg/mL fungizone (Gibco), 100 U/mL penicillin, and 100 μg/mL streptomycin. On day 35, cell aggregates were transferred to spinner-flask bioreactors (Corning) and maintained at 56 rpm, in CDM III, consisting of CDM II supplemented with 10% fetal bovine serum (FBS) (GE-Healthcare), 5 μg/mL heparin (Sigma), and 1% Matrigel (Corning). From day 70, organoids were cultured in CDM IV, consisting of CDM III supplemented with B27 supplement (Gibco) and 2% Matrigel. Please note that for these modified experiments, we eliminated the need for growth under 40% O2, the need for cell aggregates to be periodically bisected, and the use of high O_2 penetration dishes, by adapting the cultures to growth in spinner-flask bioreactors.

Neuronal culture and generation

H1 ESCs were grown to confluency, split with accutase and seeded at a density of $3{\cdot}10^5$ cells per well of a coated 6-well plate in mTeSR supplemented with 10 μM ROCK. Cells were then transduced with hNGN2 and RTTA lentiviruses to obtain >90% infection efficiency using 4 μg/mL polybrene. mTeSR was changed daily until the cells were ready to split into a single-cell suspension with accutase, and the seeded directly into N2 media, so that approximately 1.2×10^6 cells were present per 10 cm dish. After 1 day, the N2 media was replaced with N2 media supplemented with puromycin at a concentration of 0.7 ug/ml to enable selection for transduced clones, which is complemented 2 days later with B27. The media was then replaced with N2 B27 media supplemented with 2 uM Ara-C (1-β-D-Arabino-furanosylcytosine) with ½ media change every other day. Cells were then

allowed to grow and mature into neurons for 2 weeks before RNA isolation and error measurements.

Lentiviral generation and transduction

HEK293T cells (ATCC) were plated at 25% confluency and then transfected with plasmids that carry WT of or mutant versions of various proteins (GeneCopoeia, custom made plasmids) using Origene's lentiviral packaging kit (TR30037). All plasmids were custom made by Genecopoeia by Gibson assembly in their LV105, LV130, LV182 and LV183 backbone and are available upon request. Medium was replaced after 18 h of incubation and viral particles were harvested 24 and 48 h later and filtered through a 0.45 µm PES filter. The particles were then concentrated using a sucrose gradient in a Beckman ultracentrifuge at 70,000× *g* for 2.5 h at 4 °C. Afterwards, the viral pellets were resuspended in 25 µL ice cold dPBS for every 15 mL of viral medium spun down. AG10215 fibroblasts (Coriell Instititute) and U87 glioblastoma cells (ATCC) were then transduced with the concentrated viral particles at various MOIs 5 in antibiotic-free complete medium with 8 µg/mL of polybrene. Cells were incubated for 18–24 h before medium was changed to complete medium. Antibiotic selection for transduced cells began 48 h after transduction and fluorescence assessed with a Leica Stellaris confocal microscope.

Mouse neural stem cell culture

Cells were derived from mouse hippocampi, cultured at 37 °C in 5% CO_2 and 5% O_2 on PLO- and laminin-coated wells in serum-free media (NSC media) containing 1× DMEM/F12 (Invitrogen, 10565018), 1× pen/strep (Invitrogen 15140122), 1×B27 (Invitrogen, 17504044), 20 ng/ml FGF2 (PeproTech, 100-18B), 20 ng/ml EGF (PeproTech, AF-100-15) and 5 µg/mL heparin (Sigma, H3149). For quiescence induction, cells were grown for at least 3 days in the same medium as described above, but without EGFFGF2 and with the addition of 50 ng/ml BMP-4 (Fisher Scientific, 5020BP010).

MGMT protein levels

Immunoblotting: 20ug of nuclear lysates were boiled at 75 °C under denatured conditions and resolved on 4–20% gradient gels. Proteins were electroblotted using a Criterion blotter (Bio-Rad Laboratories, Hercules, CA) and transferred onto 0.45um polyvinyl difluoride membranes. Membranes were stained using Revert 700 fluorescent protein stain as a loading control and imaged prior to blocking with Intercept blocking buffer (LI-COR Biosciences, Lincoln, NE). Membranes were incubated overnight for 16 h with 1:500 MGMT primary antibody (67476-1-Ig; Proteintech, Rosemead, IL). Membranes incubated with IRDye 800CW and/or 700CW secondary antibodies and visualized with a LI-COR Odyssey C1920. Densitometry was quantified with ImageJ and normalized by total protein per lane.

Protein expression and purification

TP53 (aa 92-292) clones in Pet28a were transformed into Rosetta DE3 pLysS competent cells (Novagen) and induced by 1 mM IPTG at 18 °C overnight. Then they were purified by Ni-NTA agarose (Qiagen). After additional purification by Mono S column (GE Healthcare) and buffer exchange, they were loaded onto Superdex 75 gel filtration column (GE Healthcare) running on an ÄKTA FPLC system. *SOD1* (aa 1-154) clones in Pet28a were transformed into Rosetta DE3 pLysS competent cells (Novagen) and induced by 1 mM IPTG at 18 degree overnight. Then they were purified by Ni-NTA agarose (Qiagen), and which were further purified by Superdex 75 gel filtration column (GE Healthcare) running on an ÄKTA FPLC system

Transmission electron microscopy, atomic force microscopy, fiber growth and hanging drop method

For TEM, protein samples were spotted on carbon-coated Formvar grid (Ted Pella). The samples were stained with nanoW/uranyl acetate before air drying. The images were taken on Talos F200C G2 at 80 kV at the Core Center of Excellence in Nano Imaging (CNI). For AFM, protein samples of different seeding conditions were spotted on MICA sheets before loading on Dimension Icon(Bruker), with SCANASYST-AIR probe in ScanAsyst mode. Different dilution ratios were tested for the best visualization condition. To monitor fiber growth inside wells or by the hanging drop method, protein samples of different seeding and dilution conditions were set up either in wells or hanging drop manner. All samples were observed under polarized light to ensure fiber structure existence. A range of high concentrations of NaCl was used in the mother liquor of the hanging drop tray to induce the necessary evaporation.

Multi-angle light scattering

Experiments were conducted at the University of Southern California NanoBiophysics Core Facility. Purified WT TP53, and the $TP53^{S149F}$ mutant were subjected to HPLC chromatography Shodex KW 803 instrument, in a buffer containing 500 mM Na_2SO_4 and 10 mM acetic acid (pH 4.0). The column effluent was passed directly into a Dawn Helios MALS detector (Wyatt Technology) and an Optilab rEX refractometer (Wyatt Technology). Data was analyzed by ASTRA 6 software.

Single cell experiments

Cells were treated for 1 h (mNSCs) or 40 min (yeast) with 10µg/ml MNNG. Cells were then counted with a MacsQuant cell counter and loaded onto a 10×Genomics chip for GEM preparation according to 10x Genomics protocols, so that approximately 5000 cells would be captured inside GEMs. For yeast cells, 8000 cells were loaded with the expectation that that would result in 5000 successful GEMs as well. In addition, 1µl of zymolyase was added to the yeast cell suspension to facilitate the removal of the cell wall. Results were then analyzed by CellRanger software, and on average, 2000–9000 single cells passed QC thresholds and were successfully sequenced for each replicate.

Seurat processing for mNSC single cell RNA-seq

CellRanger output folders were imported for processing in R v3.6.3 using Seurat v3.2.2[79]. Runs from 2 independent batches were merged together for analysis. To retain only high-quality cells, we applied filters nFeature_RNA > 1000 & percent.mito <20. To determine likely cell cycle stage, a list of mouse cell cycle genes was obtained from the Seurat Vignettes (https://www.dropbox.com/s/3dby3bjsaf5arrw/cell_cycle_vignette_files.zip?dl=1), derived from a mouse study[80]. Cell cycle phase was predicted using these genes, and using the function CellCycleSorting to assign cell cycle scores to each cell. Likely Doublets were identified using DoubletFinder 2.0[81], and removed from downstream processing. Reciprocal PCA was used to integrate data from the 2 cohorts and mitigate batch effects, using the top 7500 most variable genes and with $k = 10$. To determine whether proteostasis-related terms were differentially regulated in response to DNA-damage in quiescent NSCs at the single-cell level, we leveraged the UCell robust single-cell gene signature scoring metric implemented through R package 'UCell' 1.3.1[82]. Cell-wise UCell scores were computed for selected GO terms related to proteostasis. Genes associated with these GO terms were obtained from ENSEMBL Biomart (version 109; accessed 2023-04-22) to retain relationships with all evidence codes except NAS/TAS. For analysis of statistical significance, we used ANOVA to compare the distribution of UCell scores across time points and is reported for each gene set, and p-values were corrected for multiple hypothesis testing using the Benjamini-Hochberg method.

Pseudo-allele detection

Sequencing reads are first processed with the Cell Ranger Pipeline100. For each cell, reads with the same UMI (*i.e.* PCR duplicates) are collapsed into a consensus sequence, which is incorporated into a pileup

file summarizing the sequence of each unique transcript at each genomic position in each cell. Positions covered by at least 40 unique transcripts are retained for downstream analysis and those with at least 10% of unique bases divergent from genomic DNA are compiled into a final output file for each cell.

Transfection of human cells with WT and mutant mRNAs

SOD1 mRNAs were generated from the plasmids that carry WT and mutant *SOD1*^{G142E} genes described above (tagged with *eGFP* and *mCherry* respectively) using the AMbion mMESSAGE mMACHINE T7 Ultra kit and included a 5′ ARCA cap and 3′ poly(A) tail. These molecules were then transfected into primary human fibroblasts (Nathan Shock Center UCSD) or 3T3 cells (ATCC) using he Lipofectamine MessengerMAX reagent.

Site directed RNA editing of human cells

Site directed RNA editors were generated using double stranded DNA oligonucleotides encoding the guide that were cloned into the BLOCK-iT™ U6 RNAi Entry Vector. The same approach was used to introduce the mutant versions of these guides, which were created with the Quikchange Lightning Site-directed Mutagenesis kit by Agilent Technologies[61]. Constructs contain ADAR2 linked to a gRNA using a bacteriophage lN protein BoxB hairpin linkage (IN-DD). This system contains 4 lN peptides and a nuclear localization signal added to the deaminase domain of ADAR2, as well as 2 BoxB hairpins attached to the gRNA. Constructs were transfected into cells using the Effectene® Transfection Reagent kit by QIAGEN. RNA was then extracted from cells using the RNAqeous kit from Ambion Life Technologies and processed according to standard circe-sequencing protocols[17]. Finally, the data was analyzed using a bioinformatic pipeline that called edited bases present at a frequency of either less or more than 2%.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The sequencing data that was generated have been deposited in the SRA database (https://www.ncbi.nlm.nih.gov/sra) under the following accession codes: PRJNA917136, for the H1ESC654 data, PRJNA1138749 for the mouse single cell NSC data, PRJNA673853 for human fibroblast data, PRJNA1142197 for human neuron data, PRJNA1141934 for brain organoid data, PRJNA1141225 for yeast single cell data. Source data are provided as a source data file. Source data are provided with this paper.

Code availability

Code to analyze circ-seq datasets is available at https://github.com/jfgout/circseq-seqan2. This code can also be accessed through Zenodo: https://doi.org/10.5281/zenodo.7591325.

References

1. Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
2. Hipp, M. S., Kasturi, P. & Hartl, F. U. The proteostasis network and its decline in ageing. *Nat Rev Mol Cell Biol* **20**, 421–435 (2019).
3. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188–1203 (2012).
4. Scheckel, C. & Aguzzi, A. Prions, prionoids and protein misfolding disorders. *Nature reviews. Genetics* **19**, 405–418 (2018).
5. de Oliveira, G. A. P. et al. The status of p53 oligomeric and aggregation states in cancer. *Biomolecules* **10**, 548 (2020).
6. Gertz, M. A., Dispenzieri, A. & Sher, T. Pathophysiology and treatment of cardiac amyloidosis. *Nat Rev Cardiol* **12**, 91–102 (2015).
7. Moreau, K. L. & King, J. A. Protein misfolding and aggregation in cataract disease and prospects for prevention. *Trends Mol Med*. **18**, 273–282 (2012).
8. Dember, L. M. Amyloidosis-associated kidney disease. *J Am Soc Nephrol* **17**, 3458–3471 (2006).
9. Schroder, R. Protein aggregate myopathies: the many faces of an expanding disease group. *Acta Neuropathol* **125**, 1–2 (2013).
10. Gregersen, N., Bross, P., Vang, S. & Christensen, J. H. Protein misfolding and human disease. *Annu Rev Genomics Hum Genet* **7**, 103–124 (2006).
11. Garcion, E., Wallace, B., Pelletier, L. & Wion, D. RNA mutagenesis and sporadic prion diseases. *J Theor Biol* **230**, 271–274 (2004).
12. Gout, J. F. et al. The landscape of transcription errors in eukaryotic cells. *Sci Adv* **3**, e1701484 (2017).
13. Vermulst, M. et al. Transcription errors induce proteotoxic stress and shorten cellular lifespan. *Nat Commun* **6**, 8065 (2015).
14. Saxowsky, T. T. & Doetsch, P. W. RNA polymerase encounters with DNA damage: transcription-coupled repair or transcriptional mutagenesis? *Chemical reviews* **106**, 474–488 (2006).
15. Prusiner, S. B. Scrapie prions. *Annu Rev Microbiol* **43**, 345–374 (1989).
16. Brettschneider, J., Tredici, K. Del, Lee, V. M. & Trojanowski, J. Q. Spreading of pathology in neurodegenerative diseases: a focus on human studies. *Nat Rev Neurosci* **16**, 109–120 (2015).
17. Fritsch, C., Gout, J. P. & Vermulst, M. Genome-wide surveillance of transcription errors in eukaryotic organisms. *Journal of visualized experiments: JoVE* **13**, 57731 (2018).
18. Acevedo, A., Brodsky, L. & Andino, R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* **505**, 686–690 (2014).
19. Acevedo, A. & Andino, R. Library preparation for highly accurate population sequencing of RNA viruses. *Nature protocols* **9**, 1760–1769 (2014).
20. Chung, C. et al. The fidelity of transcription in human cells. *Proc Natl Acad Sci USA* **120**, e2210038120 (2023).
21. Mead, S., Lloyd, S. & Collinge, J. Genetic factors in mammalian prion diseases. *Annu Rev Genet* **53**, 117–147 (2019).
22. Van Cauwenberghe, C., Van Broeckhoven, C. & Sleegers, K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet Med* **18**, 421–430 (2016).
23. Nguyen, H. P., Van Broeckhoven, C. & van der Zee, J. ALS genes in the genomic era and their implications for FTD. *Trends Genet* **34**, 404–423 (2018).
24. Landrum, M. J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* **46**, D1062–D1067 (2018).
25. Stenson, P. D. et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet* **136**, 665–677 (2017).
26. Sato, T. et al. Identification of two novel mutations in the Cu/Zn superoxide dismutase gene with familial amyotrophic lateral sclerosis: mass spectrometric and genomic analyses. *J Neurol Sci* **218**, 79–83 (2004).
27. Kwiatkowski, T. J. Jr. et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **323**, 1205–1208 (2009).
28. Rowe, D. B. et al. Novel prion protein gene mutation presenting with subacute PSP-like syndrome. *Neurology* **68**, 868–870 (2007).
29. Kim, M. O., Takada, L. T., Wong, K., Forner, S. A. & Geschwind, M. D. Genetic PrP Prion Diseases. *Cold Spring Harb Perspect Biol* **10**, a033134 (2018).
30. An, H. et al. ALS-linked FUS mutations confer loss and gain of function in the nucleus by promoting excessive formation of dysfunctional paraspeckles. *Acta Neuropathol Commun* **7**, 7 (2019).

31. Wang, H. et al. Mutant FUS causes DNA ligation defects to inhibit oxidative damage repair in Amyotrophic Lateral Sclerosis. *Nat Commun* **9**, 3683 (2018).
32. Ishigaki, S. & Sobue, G. Importance of Functional Loss of FUS in FTLD/ALS. *Front Mol Biosci* **5**, 44 (2018).
33. Sephton, C. F. et al. Activity-dependent FUS dysregulation disrupts synaptic homeostasis. *Proc Natl Acad Sci USA* **111**, E4769–E4778 (2014).
34. Rulten, S. L. et al. PARP-1 dependent recruitment of the amyotrophic lateral sclerosis-associated protein FUS/TLS to sites of oxidative DNA damage. *Nucleic Acids Res* **42**, 307–314 (2014).
35. Srinivasan, E. & Rajasekaran, R. A systematic and comprehensive review on disease-causing genes in amyotrophic lateral sclerosis. *J Mol Neurosci* **70**, 1742–1770 (2020).
36. Parakh, S. & Atkin, J. D. Protein folding alterations in amyotrophic lateral sclerosis. *Brain Res* **1648**, 633–649 (2016).
37. Lynch, M. et al. Genetic drift, selection and the evolution of the mutation rate. *Nature reviews. Genetics* **17**, 704–714 (2016).
38. Toombs, J. A. et al. De novo design of synthetic prion domains. *Proc Natl Acad Sci USA* **109**, 6519–6524 (2012).
39. Le Ber, I. et al. hnRNPA2B1 and hnRNPA1 mutations are rare in patients with "multisystem proteinopathy" and frontotemporal lobar degeneration phenotypes. *Neurobiol Aging* **35**, 934 e935–934 e936 (2014).
40. Kim, H. J. et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).
41. Espinosa Angarica, V. et al. PrionScan: an online database of predicted prion domains in complete proteomes. *BMC Genomics* **15**, 102 (2014).
42. Lancaster, A. K., Nutter-Upham, A., Lindquist, S. & King, O. D. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* **30**, 2501–2502 (2014).
43. Pawlicki, S., Le Bechec, A. & Delamarche, C. AMYPdb: a database dedicated to amyloid precursor proteins. *BMC Bioinformatics* **9**, 273 (2008).
44. Eneqvist, T., Andersson, K., Olofsson, A., Lundgren, E. & Sauer-Eriksson, A. E. The beta-slip: a novel concept in transthyretin amyloidosis. *Molecular cell* **6**, 1207–1218 (2000).
45. Hausser, J., Mayo, A., Keren, L. & Alon, U. Central dogma rates and the trade-off between precision and economy in gene expression. *Nat Commun* **10**, 68 (2019).
46. Fryer, H. R. & McLean, A. R. There is no safe dose of prions. *PLoS One* **6**, e23664 (2011).
47. Saxowsky, T. T., Meadows, K. L., Klungland, A. & Doetsch, P. W. 8-Oxoguanine-mediated transcriptional mutagenesis causes Ras activation in mammalian cells. *Proc Natl Acad Sci USA* **105**, 18877–18882 (2008).
48. Bregeon, D., Doddridge, Z. A., You, H. J., Weiss, B. & Doetsch, P. W. Transcriptional mutagenesis induced by uracil and 8-oxoguanine in Escherichia coli. *Molecular cell* **12**, 959–970 (2003).
49. Viswanathan, A., You, H. J. & Doetsch, P. W. Phenotypic change caused by transcriptional bypass of uracil in nondividing cells. *Science* **284**, 159–162 (1999).
50. Wyatt, M. D. & Pittman, D. L. Methylating agents and DNA repair responses: methylated bases and sources of strand breaks. *Chem Res Toxicol* **19**, 1580–1594 (2006).
51. Mu, Y. & Gage, F. H. Adult hippocampal neurogenesis and its role in Alzheimer's disease. *Mol Neurodegener* **6**, 85 (2011).
52. Teuber-Hanselmann, S., Worm, K., Macha, N. & Junker, A. MGMT-methylation in non-neoplastic diseases of the central nervous system. *Int J Mol Sci* **22**, 3845 (2021).
53. Gerson, S. L. MGMT: its role in cancer aetiology and cancer therapeutics. *Nat Rev Cancer* **4**, 296–307 (2004).
54. J. Chung et al., Genome-wide association and multi-omics studies identify MGMT as a novel risk gene for Alzheimer's disease among women. *Alzheimers Dement*, (2022).
55. Fritsch, C. et al. Genome-wide surveillance of transcription errors in response to genotoxic stress. *Proc Natl Acad Sci USA* **118**, e2004077118 (2021).
56. Chung, C. et al. Evolutionary conservation of the fidelity of transcription. *Nat Commun* **14**, 1547 (2023).
57. Konopka, A. & Atkin, J. D. DNA damage, defective DNA repair, and neurodegeneration in amyotrophic lateral sclerosis. *Front Aging Neurosci* **14**, 786420 (2022).
58. Vermulst, M. et al. MADDD-seq, a novel massively parallel sequencing tool for simultaneous detection of DNA damage and mutations. *Nucleic Acids Res* https://doi.org/10.1093/nar/gkae632 (2024).
59. Liscovitch-Brauer, N. et al. Trade-off between transcriptome plasticity and genome evolution in cephalopods. *Cell* **169**, 191–202 e111 (2017).
60. Albertin, C. B. et al. Genome and transcriptome mechanisms driving cephalopod evolution. *Nat Commun* **13**, 2427 (2022).
61. Tao, J., Bauer, D. E. & Chiarle, R. Assessing and advancing the safety of CRISPR-Cas tools: from DNA to RNA editing. *Nat Commun* **14**, 212 (2023).
62. Montiel-Gonzalez, M. F., Vallecillo-Viejo, I. C. & Rosenthal, J. J. An efficient system for selectively altering genetic information within mRNAs. *Nucleic Acids Res*. **44**, e157 (2016).
63. Montiel-Gonzalez, M. F., Vallecillo-Viejo, I., Yudowski, G. A. & Rosenthal, J. J. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc Natl Acad Sci USA* **110**, 18285–18290 (2013).
64. Falcon, B. et al. Tau filaments from multiple cases of sporadic and inherited Alzheimer's disease adopt a common fold. *Acta Neuropathol* **136**, 699–708 (2018).
65. Martinez-Vicente, M. & Cuervo, A. M. Autophagy and neurodegeneration: when the cleaning crew goes on strike. *Lancet Neurol* **6**, 352–361 (2007).
66. Tanner, C. M. et al. Rotenone, paraquat, and Parkinson's disease. *Environ Health Perspect* **119**, 866–872 (2011).
67. Spivey, A. Rotenone and paraquat linked to Parkinson's disease: human exposure study supports years of animal studies. *Environ Health Perspect* **119**, A259 (2011).
68. Spencer, P. S., Palmer, V. S. & Kisby, G. E. Western Pacific ALS-PDC: evidence implicating cycad genotoxins. *J Neurol Sci* **419**, 117185 (2020).
69. Verheijen, B. M., Hashimoto, T., Oyanagi, K. & van Leeuwen, F. W. Deposition of mutant ubiquitin in parkinsonism-dementia complex of Guam. *Acta Neuropathol Commun* **5**, 82 (2017).
70. Garruto, R. M., Yanagihara, R. & Gajdusek, D. C. Disappearance of high-incidence amyotrophic lateral sclerosis and parkinsonism-dementia on Guam. *Neurology* **35**, 193–198 (1985).
71. Verheijen, B. M. et al. The cycad genotoxin methylazoxymethanol, linked to Guam ALS/PDC, induces transcriptional mutagenesis. *Acta Neuropathol Commun* **12**, 30 (2024).
72. Farrer, L. A. et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA* **278**, 1349–1356 (1997).
73. Guo, H. H., Choe, J. & Loeb, L. A. Protein tolerance to random amino acid change. *Proc Natl Acad Sci USA* **101**, 9205–9210 (2004).
74. Debes, C. et al. Ageing-associated changes in transcriptional elongation influence longevity. *Nature* **616**, 814–821 (2023).
75. Banerjee, K., Kolomeisky, A. B. & Igoshin, O. A. Elucidating interplay of speed and accuracy in biological error correction. *Proc Natl Acad Sci USA* **114**, 5183–5188 (2017).

76. van Leeuwen, F. W. et al. Frameshift mutants of beta amyloid precursor protein and ubiquitin-B in Alzheimer's and Down patients. *Science* **279**, 242–247 (1998).
77. van Leeuwen, F. W., Burbach, J. P. & Hol, E. M. Mutations in RNA: a first example of molecular misreading in Alzheimer's disease. *Trends Neurosci* **21**, 331–335 (1998).
78. Kadoshima, T. et al. Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *Proc Natl Acad Sci USA* **110**, 20284–20289 (2013).
79. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology* **36**, 411–420 (2018).
80. Kowalczyk, M. S. et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res* **25**, 1860–1872 (2015).
81. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst* **8**, 329–337 e324 (2019).
82. Andreatta, M. & Carmona, S. J. UCell: Robust and scalable single-cell gene signature scoring. *Comput Struct Biotechnol J* **19**, 3796–3798 (2021).

Acknowledgements

Tissue for this study was obtained from the USC ADRC Neuropathology Core, NIA AG066530. The UCI ADRC is funded by NIH/NIA Grant P30 AG066519. Brain specimens were obtained from ADRC Tissue Cores: USC (P50-AG005142,AG066530); University of California Irvine (P30-AG066519); UW(P30-AG066509;U01-AG006781). M.V. was supported by NIA award R01AG054641, R01AG075130, R01AG075130 and a pilot award from SCEHSC at USC.

Author contributions

M.V., L.C. and J.G. designed the research. C.C, Y.K, S.J.S., B.V., I.F., K.L., A.D., M. T., Y.L., R.T., L.C., M.M., H.B., Y.J., A.J., D. H., Y.S., J.Q., S.S and M.V. performed experiments. D.M., K.T., M.C., Z.L., J.R., S.K., G.Q., L.C and MV oversaw and directed experiments. J.G. Y.K., B.B.and M.V. performed statistical analyzes. E.H. provided human sample materials.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-52886-2.

Correspondence and requests for materials should be addressed to Marc Vermulst.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.